L. RUBENSTEIN/BROAD INST.

David Altshuler: no one approach can explain heritability.

identified genetic variants occur at frequencies of less than 5 per cent. More than one-third of newly discovered SNPs with frequencies of less than 0.5% were observed in only one population. Such discoveries mean that many more variants can be added to microarrays for assay, and so tested in GWAS, says David Bentley, chief scientific officer at Illumina, a genetics company in San Diego, California. "There is a new generation of GWAS that are fundamentally different from previous studies, because they capture a new fraction of variations that have previously been uncharted," he says.

Illumina and other commercial vendors have been modifying their microarrays in response to releases of data. Illumina unveiled its HumanOmni2.5-Quad DNA Analysis BeadChip in June this year — letting researchers assay 2.5 million SNPs and other variants — and plans to launch the Omni5 next year, for 5 million SNPs. Using the Omni5, researchers will be able to combine one set of comprehensive SNPs with specialized sets tuned to emerging sequencing data. Illumina's competitor Affymetrix, in Santa Clara, California, has in its catalogue products geared towards Chinese, Japanese, European and African ethnicities. A new microarray design allows researchers to design custom arrays containing 50,000 up to a planned 5 million SNPs using a database stocked with proprietary and public SNP data.

Nonetheless, it is not clear how effective adding to the available SNPs from healthy populations is going to be in finding SNPs associated with disease, says Christophe Lambert, chief executive of Golden Helix, a genetic-analysis company in Bozeman, Montana. This year, his company worked on an association study for Alzheimer's disease that failed to detect a signal from a variant known to boost risk for the condition. The variant, in the gene *APOE*, wasn't included on the commercial assay used in the test. Although a custom-designed array found the variant's association with the disease to be extremely significant ($P < 10^{-60}$), the standard array did not pick up its signal. "None of the SNPs on the standard chip was correlated strongly enough with the risk variant to detect it," says Lambert. Even when Lambert's team used data from the 1000 Genomes Project to 'impute' the presence of one SNP by detecting another, the analysis did not pick up on the association. Sampling more individuals or using denser microarrays might have helped, but identifying variants in diseased individuals would produce the most-informative SNPs for genotyping across populations, says Lambert.

Still, the ability to look more deeply within populations has intriguing possibilities. In a study published this September[3], researchers at deCODE Genetics in Reykjavik found that the same SNP was associated with glaucoma risk in Chinese and Icelandic populations, but in the former it was much rarer and indicated a much higher risk. And if different susceptibility variants show up near the same gene in different populations, researchers will have independently implicated that genomic area in the disease.

DUKE UNIV.

David Goldstein: you have to choose what to pursue.

Working across populations and with rarer variants can get complicated, says Augustine Kong, head of statistics at deCODE. SNPs specific to a particular population could be difficult to replicate, and the lower the frequency of an allele, the larger the number of samples needed to detect an association. However, if rarer SNPs have stronger effects, larger sample sizes might not be necessary. Researchers are keen to find out whether a substantial number of the new variants discovered by genome-mapping projects will be associated with large effects. "Before, we just didn't have the technology to interrogate these low-frequency variants comprehensively," he says. "It gives you chances that you didn't have before to make discoveries."

SEQUENCING STRAIGHT TO CAUSAL VARIANTS

Some experts think that it is time to skip array-based GWAS that find SNPs associated with causative variants, and to hunt for contributing variants directly. Mary-Claire King is a geneticist at the University of Washington in Seattle, whose work in family studies identified the breast-cancer genes *BRCA1* and *BRCA2*. She says that even the rarer variants discovered by the 1000 Genomes Project are unlikely to be highly associated with disease. New variants

The tough new variants

When single nucleotide polymorphism (SNP) studies failed to explain much of the heritability of diseases, researchers began pinning their hopes on a trickier source of variability: copy number variation (CNV). Whereas SNPs — changes of one DNA letter into another — are relatively easy for microarrays to detect and for databases to compile and sort, CNVs are a headache to identify and classify. Certain stretches of DNA are duplicated, inverted or repeated in some individuals and missing from others. "It's more complicated and the data will always be a little more dirty," says Stephen Scherer, director of the Centre for Applied Genomics at the Hospital for Sick Children in Toronto, Canada. In some cases, researchers can detect CNVs using microarrays designed for detecting SNPs. Others use products designed to identify CNVs directly, from companies such as Agilent Technologies in Santa Clara, California, and Roche Nimblegen in Madison, Wisconsin. One Agilent array, designed with the Wellcome Trust Case Control Consortium, detects about 11,000 common CNVs.

Measuring whether a nucleotide at a particular spot is A or G is easier than detecting how many times a certain sequence occurs. That concerns Peter Donnelly, director of the Wellcome Trust Centre for Human Genetics in Oxford, UK. "Because there was a long history of GWA studies that didn't replicate, the field insists on strong criteria for declaring an association," he says. "Yet when it moves to CNVs, which are harder to measure, the standards the field requires are weaker."

The jury is out on how much CNVs matter for common diseases. A study this year[8] profiled 3,423 CNVs, or perhaps half of all those larger than 500 base pairs. It found that most not only don't explain much disease, but are also so closely associated with common SNPs that they've already been explored, albeit indirectly.

Scherer is not so sure. He was part of a team that resequenced a human genome and compared it to a reference. It found that the genome differed from the reference in only 0.1% of SNPs, but in 1.2% of CNVs. The analysis indicated that up to one-quarter of CNVs are not associated with SNPs, and so are likely to be missed by SNP studies[9].

As with SNPs, larger effects may be found in rarer and harder-to-measure variants. Scherer has done studies showing that people with autism-spectrum disorders carry more rare CNVs than do controls. To be certain that the CNVs were correctly typed, he and his colleagues ran subsets of samples through calling algorithms that convert an instrument's signals into a sequence of base pairs, and used two platforms (by Illumina, of San Diego, California, and Agilent) to identify them[10].

Scherer says that many research groups are still learning about CNVs and don't fully realize the need to validate their data. "People are looking for low-hanging fruit; they see what they want to see and publish it," he says. The situation is improving, with the maturation of databases that collect diverse data on variation. "Now that we have much better data sets to compare to, it's becoming more accurate." M.B.