

# BIO WORLD<sup>®</sup> TODAY

FRIDAY  
FEBRUARY 18, 2005

THE DAILY BIOTECHNOLOGY NEWSPAPER

VOLUME 16, No. 33  
SPECIAL REPRINT

*2005 AAAS Meeting*

## 'Matchmaker' SNP Study Paves Way For Pharmacogenomics

By Anette Breindl  
Science Editor

WASHINGTON – The annual meeting of the American Association for the Advancement of Science opened Thursday with, among other things, the presentation of new genomic data that could help shed light on the molecular basis of disease.

The work, published by researchers from Mountain View, Calif.-based biotechnology company Perlegen Sciences Inc.; the International Computer Sciences Institute in Berkeley, Calif.; and the University of California at San Diego, is published in the Feb. 18, 2005, issue of *Science*.

While genome sequence data per se are not exactly in short supply these days, there is still not much information available about genetic variation across populations and how those differences relate to disease, particularly for complex, multifactorial diseases in which one gene may only slightly influence disease risk. The overall goal of the research, which was presented at a press briefing at the AAAS annual meeting on Thursday, was to identify common genetic variants that might contribute to the molecular basis of disease and drug response.

Genes exist in different variants known as alleles; the simplest type of allele is a variation in just one base pair on the DNA sequence, or single nucleotide polymorphism, also known as SNP ("snip"). Alleles are not usually distributed 50-50, and the prevalence of the allele that occurs less frequently is known as the minor allele frequency. There are an estimated 7 million SNPs in the human genome with a minor allele frequency of at least 5 percent, and 10 million with a minor allele frequency of at least 1 percent. The researchers focused their attention on mapping such common variants, i.e., those with a minor allele frequency of at least 5 percent.

The focus on common variants represents a compromise between the useful, the desirable and the doable. David Cox, chief scientific officer at Perlegen, and colleagues

pointed out that common variants, simply because they are common, contribute to a larger proportion of disease risk and thus are more valuable from a diagnostic and therapeutic standpoint. However, they also acknowledged that investigating rare variants is experimentally more challenging.

"Detecting and characterizing effects of rare variants requires very large sample sizes to obtain statistically meaningful numbers of individuals carrying a rare allele," they wrote in the *Science* paper. "There is no doubt that rare variants play a role in the etiology of common disease, but pursuit of common variants is more tractable with available technologies."

The researchers began by identifying 2.4 million SNPs believed to be common across three populations: European Americans, African-Americans and Han Chinese Americans. They then successfully sequenced nearly 1.6 million of those SNPs in 71 people from all three populations. Most of 1.6 million SNPs were found in both allelic forms across all three populations. The scientists mapped the presence and absence of alleles as well as their frequencies across all three.

The work itself is a pure mapping study and does not attempt to correlate the identified SNPs with any clinical findings. In a press briefing, Cox, who also is senior author of the *Science* paper, described the study as "sort of a matchmaker." The idea is to "take existing treatments and identify which individuals they work in and which they don't" based on genetic information, rather than crude stratifications such as age, sex and race.

David Altshuler, director of medical and population genetics at the Whitehead Institute/MIT Center for Genome Research and author of an accompanying perspectives article in *Science*, noted at the briefing that "how predictive genetic variation will be for complex diseases is still an open question, but one that we can now answer, which is why this data is exciting."

The motto of the AAAS is “Advancing science, serving society,” and the potential social implications, or not, of the research were addressed at the press briefing as well as in an accompanying policy forum by Troy Duster, professor of sociology at New York University, in *Science*. Specifically, Cox stressed that the differences in genetic variation found across the three populations represent artifacts of the way genetic samples are categorized rather than biological validation of the notion of race.

“Our paper doesn’t have anything to do with a scientific or molecular definition of race,” he said, adding that when genetic data are collected from people across a wide geographic region, allele frequencies show gradual rather than abrupt changes. “It’s an artificial concept that human beings are categorized in groups,” he said. “The appropriate concept should be thought of as a gradient.”

### **Doing More With Fewer SNPs**

Given that over 10 million SNPs exist in the human genome, mapping them all remains a tall order. The 1.6 million SNPs mapped in the *Science* paper are substantially less, but still a lot to genotype in any one individual to determine disease associations. So another goal of the paper was to test whether the number of SNPs being genotyped could be further reduced by relying on so-called sentinel SNPs.

Because DNA recombines in stretches, SNPs located near each other tend to be inherited together, a process known as linkage disequilibrium. Based on the principles of linkage disequilibrium, the genome can be parsed into longer stretches known as haplotypes, and haplotype mapping of the human genome is ongoing. (See *BioWorld Today*, Oct. 30, 2002.)

In the research published here, Cox and his colleagues developed a similar modeling technique known as binning to identify sentinel SNPs. While haplotypes are defined as continuous stretches of DNA, DNA sequences from two different bins can be interspersed with each other. Based on the results of their binning analysis, the scientist estimated that on the order of a few hundred thousand SNPs should be sufficient to capture most common variants of the human genome. Whether that’s a little or a lot depends on whether you compare it to 10 million, and whether you are the person that has to do the genotyping. ■