

GENOMICS

TACLing rare genetic variants

By combining methods for selective genome capture, allele enrichment and array resequencing, researchers create a pipeline for high-throughput variant detection.

Genetic variation is the basis for the complex traits that make each of us unique and interesting. It is also the basis for many diseases, and understanding the link between variants and a specific illness has been the object of many genome-wide association studies. Malek Faham from the molecular diagnostics company MLC Dx sees both strength and weaknesses in these studies: “There have been a lot of large-scale association studies for common [single-nucleotide polymorphisms] on array platforms,” he says. “Many discoveries have been made, but they only explain a small fraction of all the genetic variation.” A lot of variants, mainly rare alleles, are yet to be discovered.

Despite their rare occurrence, these alleles often pose a large relative risk for the individual who carries them and are thus of biological interest. To find infrequent variants, Faham needed to resequence large numbers of samples, something current next-generation sequencing platforms have not yet been shown to do at a reasonable cost. DNA resequencing arrays are theoretically much better suited for this purpose but have never been used for the resequencing of large sample sizes.

Faham wanted to know why not and came across two obstacles that prevented the application of arrays to high-throughput resequencing. The first was that upstream target preparation techniques to amplify specific genomic targets had not been multiplexed to the degree that would allow them to be used on thousands of samples simultaneously. The second was that accuracy was too low. To find rare variants with a very low a priori probability, Faham was looking for a false positive rate of 10^{-5} or lower.

Together with a research team from Genentech and Ronald Davis from the Stanford Genome Technology Center, Faham, then working at Affymetrix, solved these two problems. First they developed a method for target amplification by capture and ligation (TACL) that allowed them to simultaneously target 10,000 different loci for amplification. They created

short probes from genomic DNA, initially by individually amplifying each target by PCR, later by synthesis of the probes on an array. Then they incorporated deoxyuridine into these probes and used them to capture genomic targets from the samples of interest. In the course of purification of the captured probes the deoxyuridine strand was removed again. The researchers showed that TACL provided high reproducibility and specificity in terms of capturing the targeted regions and was effective even at starting sample concentrations as low as 15 nanograms.

“That solved the first problem of target prep,” Faham concludes. “The second was accuracy. To improve accuracy we needed to do allelic enrichment.” To separate the captured targets that contained variants from those that did not, the researchers used mismatch repair detection. They cloned the pooled TACL probes into bacteria and used them to hybridize to probes captured in the previous TACL reaction. The researchers then transformed these hybrids into bacteria. If there was a mismatch, the bacteria would grow in one kind of medium, if there was none, they would grow in another. This left the scientists with homogenous variant and non-variant pools that they then resequenced on arrays.

Faham and his colleagues targeted 1,500 genes, 5 megabytes of sequence, in each of just under 500 samples and used 24 HapMap samples to determine the false positive rate of 1 in 500,000 base pairs. Of the unique variants they detected, 80% were novel.

The potential of this approach is not only in its different applications—Faham’s main interest is in finding rare genetic variants, whereas his colleagues at Genentech are exploring somatic mutations in cancer—but also in its modular nature. The three key elements, target capture, allele enrichment and resequencing, can be combined or exchanged with more current technology.

It is time to mix and match for genomic population studies.

Nicole Rusk

RESEARCH PAPERS

Zheng, J. *et al.* High-throughput, high-accuracy array-based resequencing. *Proc. Natl. Acad. Sci. USA* **106**, 6712–6717 (2009).