GENOMICS

One genome, two haplotypes

Two approaches using either fosmid clones or a microfluidic device are used to tackle the challenge of a haplotype-resolved human genome.

Although sequencing the six gigabases of the human genome has become almost trivial, sequencing the two times three gigabases of the human genome is anything but. High-throughput sequencing technologies, which allow genome sequencing in a matter of days, apply a shotgun approach that can find the polymorphisms on homologous chromosomes but cannot determine the haplotypes; that is, they cannot tell which polymorphisms lie together on the same chromosome.

Haplotype information is important to answer various questions. For example, whereas a single polymorphism in a gene is harmless, the combination of two is not; thus one needs to know whether these polymorphisms occur on the same allele.

Although several methods for haplotyping single loci exist, there is a dearth of methods for haplotyping on a genome-wide scale. If the genomes of multiple family members are sequenced, one can infer the haplotype of a closely related individual, or one might do population-based inference if genotype information on a large cohort is available. But these methods do not apply to isolated individuals.

Two independent groups have recently tackled this problem. Jay Shendure and colleagues from the University of Washington, Seattle sequenced haplotype-resolved genomes in 400-kilobase increments (Kitzman *et al.*, 2010), and Stephen Quake and co-workers from Stanford University genotyped each haploid chromosome from a single cell (Fan *et al.*, 2010).

Shendure and colleagues started with a fosmid library of genomic DNA of an Indian individual with an insert size of 40 kilobases and split it into 115 pools of ~5,000 clones each so that each pool covered about 3% of the genome. After sequencing the pools, which were derived to 99% from one of two homologous

Image of Quake and colleagues' microfluidic device to partition chromosomes. Reprinted from *Nature Biotechnology*.

chromosomes, they could define 400-kilobase haplotypes, which correlated to 99.9% with HapMap findings.

Two observations stood out for Shendure: one was the ability to “do very basic population genetics,” as he called it. The researchers saw that haplotypes, not previously seen in HapMap or the 1000 Genomes data, were the most enriched for novel variations. “This is something that is highly expected based on what we know,” says Shendure, “but seeing it empirically was fun.” The other serendipitous finding was an improvement of *de novo* genome assembly. Recent shotgun sequencing data of individual genomes have brought to light sequences that are not present in the reference genome and therefore cannot be mapped to a chromosome. Shendure and colleagues aligned the unmapped reads from their clones to these contigs, then traced them back to the fosmid pools they came from and looked for a shared location in these fosmids. In several instances this allowed them to anchor previously unmapped contigs.

Though the resolution obtained with the fosmid pool approach is high, the 400-kilobase haplotypes cannot be combined into a whole chromosome.

To achieve the goal of haplotyping an entire chromosome, Quake and colleagues developed a microfluidics device that physically separates the chromosomes of a single cell in metaphase into 48 distinct channels, allowing amplification of the DNA in the channels and retrieval of the material for genotyping

on high-density single-nucleotide polymorphism arrays. The researchers presented the individual haplotypes of four people, which allowed them, for example, to directly determine the human leukocyte antigen haplotypes of individuals.

One of the main differences between the two methods is the way the DNA is amplified. Shendure's team used bacteria to amplify the fosmid clones, but Quake and colleagues used *in vitro* multiple displacement amplification (MDA). Shendure thinks that although *in vitro* methods are easier to perform, they introduce substantial bias in coverage. Quake's team analyzed the bias introduced by MDA on chromosome 6. They saw hotspots with strong coverage bias, but Quake does not see this as a hindrance to haplotype sequencing. “If you sequence a haploid genome,” he says, “you don't need the 30–40× coverage you need for diploid base calls. We need much less coverage if we are calling just one base instead of two. We are able to cover as much of the chromosome as we want without ridiculously deep sequencing.” They have not shown sequencing of a whole haploid genome, but Quake's prediction is that “it will be easy to sequence genomes in a haploid fashion, bias notwithstanding.”

Both groups are working on improving their approaches. Quake aims to show that the current partitioning and amplification method is indeed sufficient not only for genotyping but also for sequencing. Shendure is working on an *in vitro* amplification approach that would obviate the need for cloning while at the same time not introducing the bias of current MDA methods.

It is likely that in the near future a haplotype-resolved genome will be the norm rather than the exception.

Nicole Rusk

RESEARCH PAPERS

Fan, H. C. *et al.* Whole-genome molecular haplotyping of single cells. *Nat. Biotechnol.* **29**, 51–57 (2011).

Kitzman, J.O. *et al.* Haplotype-resolved genome sequencing of a Gujarati Indian individual. *Nat. Biotechnol.* **29**, 59–63 (2011).