GENOMICS

Putting long-range mapping in reach

A technique for performing PCR on extremely large DNA molecules offers researchers the ability to generate accurate multi-marker haplotypes for large numbers of patients—even with loci arrayed across entire human chromosomes.

Haplotyping—analysis of the allelic composition of individual chromosomes—is a hot topic these days, with the promise of yielding data of great interest to researchers in both academia and industry. Among other things, patterns of association between genomic polymorphisms can yield important information about loci that may be associated with heritable diseases, or that may contribute to differences in individual responses to different drugs. Several useful techniques have been developed to date, but none of them allow truly large-scale multiplexing—the simultaneous sequencing of multiple markers from multiple patients, using loci scattered over large distances.

Early multiplex approaches, based on the probing of DNA bands that had been transferred from a gel to a solid membrane, proved to be of limited value for genomic sequence analysis. Several years ago, however, Harvard Medical School investigator George Church and his colleagues came up with an innovative alternative. “I thought we could bring up the signal-to-noise [ratio] if we could amplify each of those bands... [and] it seemed like the most obvious thing would be to do a PCR *in situ*,” he says. The result was PCR colonies, or ‘polonies’, generated by *in situ* amplification of dispersed single DNA molecules immobilized on a glass slide within a layer of polymerized acrylamide (Mitra & Church, 1999). Polonies proved to be a robust platform for sequence analysis. “Rather than having a separate cloning step, you basically put a single molecule and you grow it up to a big mass,” says Church. “You just keep the whole thing *in vitro*, and go straight from a library to sequencing all in one place, on a slide.”

Having previously demonstrated the effectiveness of polonies for small-scale haplotype

Figure 1 | Haplotyping of two markers on human chromosome 7. The top portion shows the chromosome, with colored nucleotides indicating the SNP being typed and the color used for a given sequence variant. The lower portion shows actual typing results, with intermediate colors (green and white) indicating polonies for which two SNPs have been successfully haplotyped. Reprinted with premission from *Nature Genetics*.

analysis, Church’s team has now refined their methods of DNA isolation and preparation to an extent where polonies can be derived from entire, intact human chromosomes (Zhang *et al.*, 2006). These can then be subjected to the same sort of analysis that has been previously restricted to far smaller fragments. In terms of multiplexing this method, Church gives much credit to lead author Kun Zhang: “Kun reckoned that with the same effort, [as a small-scale haplotyping experiment] you could [analyze] multiple people and multiple loci. So each polony would have a high density of information, and you’d have a high density of polonies.”

For polony haplotyping, PCR is performed with primer pairs flanking each variable locus of interest, resulting in simultaneous amplification of multiple single-nucleotide polymorphisms (SNPs). These can then be visually queried by single-base extension, which uses fluorescent nucleotides to reveal which base is present at a given variable site for each polony. This process can then be repeated, allowing researchers to potentially analyze dozens of SNPs with a single slide.

The method is fairly straightforward and cost-effective, and genomic distance appears

to be no obstacle, as Church’s team demonstrated by haplotyping five widely dispersed SNPs on human chromosome 7, spanning a distance of nearly 160 megabases (Fig. 1). They then tested the technique using samples pooled from groups of patients, generating haplotypes for 14 SNPs within the CD36 gene of 24 individuals, then comparing the resulting linkage data against the results generated by various algorithms that infer the occurrence of specific haplotypes based on statistical information. The accuracy of the polony strategy generally surpassed that of the inference-based algorithms, especially for rare haplotypes. According to Church, this is unsurprising, owing to the inherent advantages of direct sequencing methods versus computational prediction. “If you just have an isolated individual being haplotyped, then you’re just going to be guessing based on frequencies in populations and so forth, and you could be wrong,” he explains. “Especially if you’re looking at point mutations that are relatively uncommon.”

Church downplays the relative value of conventional strategies for using haplotype analysis to predict disease by assessing linkage to specific SNPs. “The major push of our lab is to try to basically sequence either whole genomes, or... the regions that *a priori* are more likely to mutate into a causative allele.” The relevance of haplotyping would then be to identify pertinent associations between these alleles. “If I have all of the nucleotide changes in my genome, or all the important ones, I still would need to know who’s on the same strand,” he says. “If you focus on the causality rather than the linkage aspect, which is ultimately where I think we’re headed... that’s where haplotypes are important.”

Michael Eisenstein

RESEARCH PAPERS

Zhang, K. *et al.* Long-range polony haplotyping of individual human chromosome molecules. *Nat. Genet.* **38**, 382–387 (2006).

Mitra, R.D. & Church, G.M. *In situ* localized amplification and contact replication of many individual DNA molecules. *Nucleic Acids Res.* **27**, e34 (1999).