Invited Lecture:

NEW APPROACHES TO GENOMIC ANALAYSIS USING SINGLE MOLECULES

Schwartz[1,2], D.C., Anantharaman[2], T., Aston[1], C., Mishra[2], B., Clarke[1], V., Gebauer[1], D., Delobette[1], S., Dimalanta[1], E., Edington[1], J., Evenzehav[1], A., Giacalone[1], J., Hiort[1], C., Huff[1], E., Jing[1], J., Lai[1], Z., Lee[1], E., Lin, J., Porter[1], B., Qi[1], R., Skiadis, Y.

1 W. M. Keck Laboratory for Biomolecular Imaging, Department of Chemistry
New York University
31 Washington Place
New York, NY 10003

2 Courant Institute of Mathematical Sciences, Department of Computer Science
New York University
251 Mercer Street
New York, NY 100012

Abstract

Current molecular biology techniques were developed primarily for characterization of single genes, not entire genomes, and, as such, are not ideally suited to high resolution analysis of complex traits and the molecular genetics of very large populations. Despite rapid progress in the human genome project effort, there is little doubt that radically new conceptual approaches are needed before routine whole genome-based analyses can be undertaken by both basic research and clinical laboratories.

Physical mapping of genomes, using restriction endonucleases, has played a major role in the identification and characterizing various loci, for example, by aiding clone contig formation and by characterizing genetic lesions. Restriction maps provide precise genomic distances, unlike ordered sequence-based landmarks such as Sequence Tagged Sites (STSs), that are essential for optimizing the efficiency of sequencing efforts, and for determining the spatial relationships of specific loci. When compared to tedious hybridization-based fingerprinting approaches, ordered restriction maps offer relatively unambiguous clone characterization that is useful in contig formation, establishment of minimal tiling paths for sequencing, and preliminary characterization of sequence lesions. In addition, such maps provide a useful scaffold for sequence assembly, often critical in the final sequence finishing stage. Despite the broad applications of restriction maps, the associated techniques for their generation have changed little over the last ten years, primarily because they still utilize electrophoretic analysis. To help overcome these shortcomings, our laboratory developed the first practical non-electrophoretic genomic mapping approach, Optical Mapping, to meet this need.

Optical Mapping is a single molecule methodology for the rapid production of ordered restriction

maps from single DNA molecules. Ordered restriction maps were constructed originally from yeast chromosomes by imaging restriction endonuclease cutting events on single, stained DNA molecules with fluorescence microscopy. Cut sites appeared as gaps that widened as the DNA fragments relaxed. Maps were then constructed by measuring fragment sizes via relative fluorescence intensity or apparent length measurements. Modern Optical Mapping technology uses aminosilane treated surfaces to adhere molecules prior to digestion. In this way, multiple samples can be robotically gridded onto a single surface and digested in parallel. Deposition techniques developed in our laboratory elongate and fix molecules to these surfaces, while retaining biochemical accessibility of samples. Following staining with a fluorochrome, cleaved molecules are imaged by a fully-automated microscope system, developed in our laboratory. Importantly, cleaved molecule fragments retain their order, facilitating fragment sizing and obviating complicated schemes to re-establish fragment order. The final map is of course, a very informative ordered restriction map instead of a mere fingerprint.

Intensive effort in our laboratory have been directed to the development of machine vision systems, and map construction algorithms to automatically construct maps from images of digested molecules. This analysis is based on Bayesian inference techniques and enables the construction of maps from noisy data. For example, the map construction algorithms can produce maps from a population of partially digested clone molecules, (BACs, cosmids, phage) having a digestion rate as low as 15%.

Using the approaches discussed above, this laboratory has generated ordered restriction maps for the Beckwith-Wiedeman locus in humans (in collaboration with Dr. D. Housman's group at MIT), the Brca2 locus (in collaboration with Dr. S. Fisher's group at Columbia University), and the mouse olfactory locus (in collaboration with Dr. R. Axel's group at Columbia University). Optical Maps are currently being generated from phage, cosmid, YAC and Bacterial Artificial Chromosome (BAC) clones. Our laboratory has been extensively mapping BAC contigs derived from the human Y chromosome, in collaboration with Dr. David Page's group at MIT. The aims are to disambiguate clones and markers to provide the basis from which to critically understand the functionality of many loci, and provide a scaffold for sequence assembly. The detailed analysis of the Y chromosome requires such detailed maps, since it is highly punctuated with repeated sequences, that frequently confound traditional techniques of characterization.

Recent efforts have been directed at the high resolution mapping of bacterial (in collaboration with Dr. O. White, TIGR) and parasite genomes (in collaboration with Drs. M. Gardner, D. Carucci; TIGR, Naval Medical Research Institute) to provide high resolution scaffolds for facilitated sequence assembly and verification. Here, we have been using high molecular weight DNA gently extracted from cells, entirely obviating the need for the mapping of clonal material. We have used Optical Mapping to generate physical maps of two microbial genomes. Nhe I maps of the E. coli (4.6Mb) and Deinococcus radiodurans (3.1Mb) genomes were generated from chromosomal DNA, obviating the use of clones for the construction of primary maps. DNA samples, prepared from gel inserts, were fixed onto derivatized glass surfaces and molecules as large as 2.4 Mb were measured. Co-mounted lambda bacteriophage DNA was used as a sizing standard and to estimate cutting efficiency. Contig maps were created by aligning maps from multiple molecules. To benchmark our system, we compared the E. coli Nhe I optical map with the map predicted from the published sequence. The 150 fragment optical map had average fragment size 30 kb and a relative sizing error of 5 per cent for fragments ¿5 kb. We then generated a whole genome Nhe I map of the D. radiodurans genome. The final map was assembled without gaps at an average depth of 35, using 157 molecules with an average restriction fragment size of 29 kb. This map will significantly aid sequence assembly and verification to collaborators at TIGR, sequencing this microorganism.

Given the success we enjoyed in the restriction mapping of whole microbial genomes, and the proven reliability of the contig assembly algorithms developed for these efforts, we decided to construct a reference restriction map of the entire human genome. In four weeks our laboratory mapped 0.6 human genome equivalents at 40 kb resolution, using genomic fragments with average size of 2.1 mb. Our analysis of the contigs formed showed good correspondence with suitably modified Lander-Waterman physical mapping criteria in terms of the number and depth of overlapped genomic fragments. Goals are to simultaneously complete the human reference map to include 10-15x coverage and to link with other physical maps by the alignment of restriction mapped BAC contigs.

Full automation of Optical Mapping holds enormous promise for miniaturization, with expected increases in throughput and reductions in cost. Thus, advantages of Optical Mapping include high throughput and resolution, safety, and low cost. Compared with traditional electrophoresis-based methods, Optical Mapping produces information rich physical maps for whole genomes with much higher a resolution. High throughput and the obviation of clones makes Optical Mapping ideally suited for population-based genomic studies.We expect that the advantages of Optical Mapping will facilitate closure of the initial objectives of the human genome project, and aid in reducing costs associated with the sequencing of microbial genomes.