Available online at www.sciencedirect.com

SCIENCE DIRECT®

Genomics 85 (2005) 1–15

GENOMICS

www.elsevier.com/locate/ygeno

Review

Applications of DNA tiling arrays for whole-genome analysis

Todd C. Mockler[a], Joseph R. Ecker[a,b,*]

[a]*Plant Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92037, USA*
[b]*Genomic Analysis Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92037, USA*

Received 7 July 2004; accepted 14 October 2004
Available online 24 November 2004

Abstract

DNA microarrays are a well-established technology for measuring gene expression levels. Microarrays designed for this purpose use relatively few probes for each gene and are biased toward known and predicted gene structures. Recently, high-density oligonucleotide-based whole-genome microarrays have emerged as a preferred platform for genomic analysis beyond simple gene expression profiling. Potential uses for such whole-genome arrays include empirical annotation of the transcriptome, chromatin-immunoprecipitation-chip studies, analysis of alternative splicing, characterization of the methylome (the methylation state of the genome), polymorphism discovery and genotyping, comparative genome hybridization, and genome resequencing. Here we review different whole-genome microarray designs and applications of this technology to obtain a wide variety of genomic scale information.

Contents

The completion of numerous genome sequences has introduced an era of whole-genome study. Gaining a more complete understanding of the genome's information content will dramatically improve our understanding of various biological processes. In parallel with the sequencing of entire genomes, recent advances in microarray technologies have made it feasible to interrogate an entire genome sequence with arrays. Such high-density whole-genome DNA microarrays can be used as a generic platform for numerous experimental approaches to decode the information contained within the genome. In this review, we discuss several approaches using high-density whole-genome oligonucleotide microarrays for transcriptome characterization, novel gene discovery, analysis of alternative splicing, mapping of regulatory DNA motifs using the chromatin-

* Corresponding author. Plant Biology Laboratory, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Rd., La Jolla, CA 92037, United States. Fax: +1 858 558 6379.
E-mail address: ecker@salk.edu (J.R. Ecker).

0888-7543/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2004.10.005

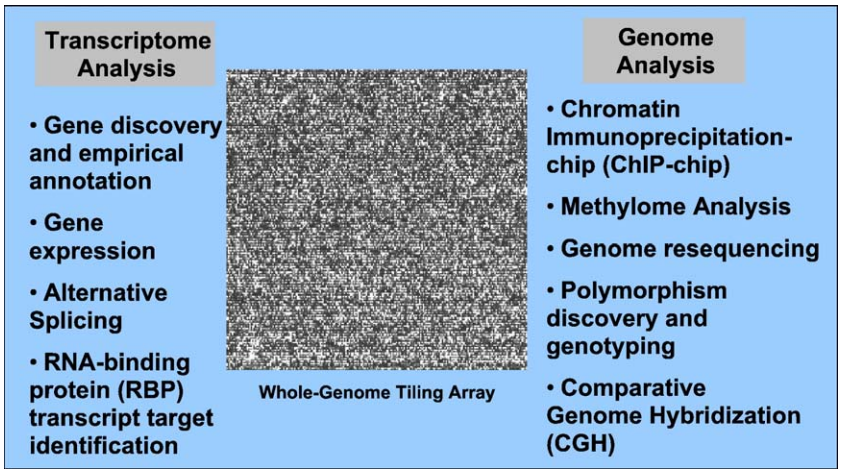

Fig. 1. Whole-genome high-density tiling arrays provide a universal data capture platform for a variety of genomic information.

immunoprecipitation (ChIP) chip, whole-genome DNA methylation analysis, polymorphism analysis, comparative genome analysis, and genome resequencing (Fig. 1).

There are two general types of high-density microarray platforms that are most widely used. High-density oligonucleotide arrays contain relatively short (<100-mer) probes synthesized directly on the surface of the arrays by photolithography using light-sensitive synthetic chemistry and photolithographic masks [1–3], an ink-jet device [4], or programmable optical mirrors [5–7]. Oligonucleotide arrays can be made with >6,000,000 discrete features per chip, with each feature comprising millions of copies of a distinct probe sequence. The second array platform is made by mechanically printing/spotting probes, generally amplified PCR products, oligonucleotides, or cloned DNA fragments, onto glass slides. This type of array generally has a much lower feature density than the in situ synthesized oligonucleotide arrays, typically of about 10,000–40,000 spots per chip.

Because oligonucleotide arrays offer a much higher feature density and high reproducibility, and probes can be synthesized to represent virtually any sequence of a finished genome, they are the preferred platform for whole-genome analysis. Moreover, the relatively short probe length combined with the flexibility of using multiple overlapping probes representing the same genomic region makes oligonucleotide arrays ideal for detecting the broadest range of genomic features, including small polymorphisms and splice variants, and the specificity also potentially allows repetitive regions or gene family members to be distinguished.

For the purpose of this review, we will distinguish whole-genome tiling arrays designed to interrogate an entire genome in an unbiased fashion [8–11] from quasi-whole-genome (nontiling) arrays, or whole-genome expression arrays, that represent the known and predicted (annotated) features of a genome, such as exons or splice junctions on a whole-genome scale [12,13]. Nonoverlapping or partially overlapping probes (Fig. 2A) may be tiled to cover the entire genome end to end, or the probes may be spaced at regular intervals (Fig. 2B). Such an unbiased approach allows researchers to analyze various features of the genome, including evidence of transcriptional activity, binding of transcriptional regulators, and DNA methylation, at high resolution without reference to prior annotations. Other array designs rely on prior genome annotation to interrogate a particular subset of features of an entire genome (Fig. 2C). These arrays are clearly limited by the quality and completeness of the annotations on which they are based.

Applications of whole-genome arrays (WGAs)

Unbiased measure of transcriptional activity

It is well known that having a finished genome sequence is not sufficient to identify all of the transcription units, as computational gene prediction methodologies are fraught with errors. While traditional molecular approaches to identifying genes, including cloning and sequencing large collections of cDNAs, have succeeded at identifying expressed transcripts for tens of thousands of genes [11,14–17], they eventually reach a point of greatly diminished returns. Transcripts that are low abundance or expressed in rare cell types or in response to specific stimuli may never be identified by these methods. Microarrays can be used to circumvent some of these problems, allowing confirmation of the predicted gene models as well as being a tool for new gene discovery. One study [13] used ink-jet-fabricated oligonucleotide arrays to study gene transcription and transcript structure on human chromosome 22 and the human genome as a whole. Exon-scanning arrays containing relatively long (50- to 60-mer) oligonucleotide probes were hybridized with labeled cDNAs derived from various cell lines and tissues and the resulting probe intensities were used to identify expressed exons, model gene structures, and compare differential expression across conditions. Overall, expression was detected for ~57% of Genscan-predicted genes; however, this study did not interrogate most of the nonrepetitive sequence on chromosome 22. Because the