Genomics 96 (2010) 308–315

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Methods

Development of a versatile, target-oriented tiling microarray assay for measuring allele-specific gene expression

Hang He [a,b,1], Huiyong Zhang [b,1], Xiangfeng Wang [c], Nicholas Wu [d], Xiaozeng Yang [d], Runsheng Chen [a], Yi Li [b], Xing Wang Deng [c,*], Lei Li [d,*]

[a] Bioinformatics Laboratory, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China
[b] Peking-Yale Joint Center of Plant Molecular Genetics and Agrobiotechnology, National Laboratory of Protein Engineering and Plant Genetic Engineering, College of Life Sciences, Peking University, Beijing 100871, China
[c] Department of MCD Biology, Yale University, New Haven, CT 06520, USA
[d] Department of Biology, University of Virginia, Charlottesville, VA 22904, USA

ARTICLE INFO

Article history:
Received 26 May 2010
Accepted 28 July 2010
Available online 3 August 2010

Keywords:
Gene expression
Tiling microarray
Allele
Hybrid
Rice

ABSTRACT

In the study of gene expression, it is often desirable to distinguish transcript pools derived from different alleles present in the same organism. We report here an oligonucleotide tiling microarray designed to specifically target 518 single nucleotide polymorphisms (SNPs) between the two sequenced rice (*Oryza sativa*) subspecies *indica* and *japonica*. The tiling array included all 25-mer probes interrogating each SNP by placing the polymorphic site at all 25 possible positions within the probe. Through hybridization to a titration series in which the *japonica*- and *indica*-derived cDNA templates were mixed with altering proportions, a regression model was used to screen for diagnostic probe sets for each SNP. Our result indicates that 284 (55%) SNPs have at least one diagnostic probe pair suitable for distinguishing and quantifying the relative abundance of allele-specific transcripts. As a proof-of-concept, we analyzed allele-specific expression in reciprocal *indica*×*japonica* F_1 hybrids and detected imbalanced expression at approximately one third of the SNPs. These results were validated by RNA-sequencing and allele-specific real-time PCR experiments. Together, our work demonstrates the utility and advantages of the tiling array method in interrogating large numbers of SNPs for quantifying allele-specific gene expression.

1. Introduction

Elucidation of the changes in gene expression associated with biological processes and developmental programs and understanding the underlying mechanisms have been a central theme in biology. Advances in molecular and computational biology in recent years have led to the development or improvement of methods for analyzing global gene expression with ever increasing experimental throughput [1,2], transcriptome coverage [3–5], and cellular resolution [6,7]. In most of these efforts, it is assumed that alleles of different origins contribute equally to the transcript pool and hence only the sum was measured. However, without allele-specific information, interpretation of gene expression could be complicated by allele-specific variation (*cis* effect), variation in other regulatory genes (*trans* effect), as well as environmental influences [8–10]. Results generated under this experimental setting thus do not offer many mechanistic cues regarding the complex allelic interactions that occur when more than one set of alleles are present in the same cell.

In diploid eukaryotic organisms, it is clear that many genes are not equally expressed from the paternal and maternal chromosomes. At the extreme are the imprinted genes that are exclusively transcribed from the non-silenced parental chromosome. We use the term "imbalanced allelic expression" (IAE) hereafter to describe variation in gene expression where alleles of the same gene are not expressed equally at the mRNA level. IAE appears to be common in heterozygous individuals [11]. For example, studies in human revealed that up to 50% of the investigated genes may exhibit IAE in heterozygote [12–14]. In plants, investigations of the transcript levels of small sets of genes indicate that IAE is potentially prevalent in heterotic F_1 hybrids [10,15,16]. These results provide a useful readout for pinpointing regulatory polymorphisms residing on the same DNA molecule that are important for controlling proper gene expression.

In angiosperm plants, polyploidy has been a prominent force in the evolution of genome organization and gene expression regulation [17]. In particular, recently formed allopolyploids typically retain duplicated copies of most genes on homeologous chromosomes that share a very high degree of sequence similarity. Numerous studies in polyploid plant species on subsets of the genome indicate that unequal expression of the homologous alleles is quite common [18–20]. Therefore, the experimental capacity to discern the genomic origin of expressed homologous

* Corresponding authors.
E-mail addresses: xingwang.deng@yale.edu (X. Deng), ll4jn@virginia.edu (L. Li).
[1] These authors had equal contribution.

0888-7543/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2010.07.008

genes and quantify their regulated contribution to the transcript pool at the genome scale is much desired. Such capacity should facilitate various studies aimed at elucidating homologous gene regulation and the impact of polyploidy on genome evolution.

SNP is the most abundant form of DNA polymorphism and serves as a valuable molecular marker for genetic studies. Not surprisingly, much of the effort to experimentally distinguish the transcript of one allele from its highly similar counterpart has been directed toward SNPs in the coding regions. Experimental procedures based on several different principles to examine unequal transcription from SNP-defined alleles were successfully developed. One type of such methods employs the physical properties of the complementary DNA molecule such as mass [15] or DNA melting [21]. Another type of methods relies on enzymatic reactions that have different efficiency at the polymorphic sites. These include, for example, the RNase protection assay [22] and single nucleotide primer extension [12,23]. Microarray-based applications including mini-sequencing on microarray [24] and allele-specific microarray [14,25,26] were also developed.

Allele-specific microarrays typically involve pairs of probes with perfect match to one of the alleles. They provide a high throughput, multiplex, target-oriented platform for globally quantifying IAE that is affordable to most research laboratories. The power of this approach lies in the idea that a sequence mismatch between a probe and its target may significantly disrupt hybridization and attenuate that probe's signal. In developing genome scale allele-specific microarrays, a key technical consideration obviously is to design probes that offer sufficient sensitivity and specificity in discriminating the transcripts derived from different alleles. An equally critical yet less addressed consideration is that quantitative measurement of IAE requires the relative hybridization signal from the allele-specific probes to respond linearly to changes in the allele-specific transcript level. Because both the actual polymorphism and its sequence context will impact probe-target hybridization, a computation-based approach to select informative SNPs to discriminate expressed alleles remains insufficient. This may prove to be the major rate limiting step in the application of allele-specific microarrays in various species.

Advancement in high-density microarray technology permitted the development of tiling microarray that involves the representation of a genomic region with progressive oligonucleotide probes. Tiling arrays have been widely used in transcriptomic studies in plants [3,4,27–31]. Here we investigate the potential of tiling array in detecting and quantifying IAE focusing on 518 semi-randomly selected SNPs between two rice subspecies *indica* and *japonica*. Our effort revealed that over half of the SNPs are diagnostic and generate accurate IAE measurement in the reciprocal *japonica*×*indica* hybrids. Further analysis indicates that the tiling microarray-based experimental approach offers a versatile, target-oriented assay for examining IAE that can be readily applied in species for which DNA polymorphism information is available.

2. Results

2.1. Design of SNP-based tiling microarray

We chose rice to design allele-discriminating tiling microarray as both *indica* and *japonica* are completely sequenced and abundantly high quality SNPs identified [32–34]. In this study, we selected 518 SNPs between *indica* and *japonica* in the coding region of 475 genes (Table 1). To further verify the quality of these SNPs, we chose 21 that were predicted to result in the creation or disruption of a restriction site and performed cleaved amplified polymorphic sequence analysis. All 21 loci were verified in this analysis (Fig. S1), indicating that the annotated SNPs are a reliable source for designing probes specific to the *indica* or *japonica* alleles.

The schematic representation of the tiling strategy is illustrated in Fig. 1. The tiling design involves 25 sets (blocks) of 25-mer probes for each of the 518 SNPs. Within each block, which is used as a unit for measuring allelic output, there are two probes that match perfectly to the *indica* and *japonica* alleles, respectively. The other two probes, called mismatch probe, each contains one of the two remaining nucleotides at the SNP site. For each SNP, the 25 blocks differ in the position of the polymorphic sites, which were placed at all 25 possible positions within the probe (Fig. 1A). Thus all possible 25-mer probes spanning the polymorphic site for each SNP were included in the tiling design. The resultant 51,800 (518 times 100) probes were synthesized in triplicate (155,400 in total) in a single microarray, which was used throughout this study.

Table 1
Number of probe blocks, SNPs, and genes in the tiling microarray analysis.

	All	Diagnostic	IAE detected[a]	
			Japonica×*indica*	*Indica*×*japonica*
Blocks	12,590	519	115	85
SNPs	518	284	95	71
Genes	475	271	93	70

[a] In the F_1 hybrids. *japonica*×*indica*, *japonica* (♀)×*indica* (♂) F_1; and *indica*×*japonica*, *indica* (♀)×*japonica* (♂) F_1.

2.2. Identification of diagnostic blocks

To experimentally identify diagnostic probe blocks, we prepared a titration series consisting of five cDNA mixtures in which the *japonica* and *indica*-derived cDNA templates are mixed with the following proportions: 1:4, 1:2, 1:1, 2:1, and 4:1. We hybridized the SNP-based tiling microarray to this titration series and obtained five sets of hybridization data (Fig. 1B). For each block, we then calculated the relative abundance of the *indica*- and *japonica*-specific transcript (i/j) using the mismatch probes as background (see Methods). To score the 12,950 (528×25) blocks, we calculated the correlation coefficient (r) between the measured i/j ratio and the known *indica*/*japonica* ratio in the input cDNA across the titration series (Fig. 1C).

Individually measured i/j ratios do not necessarily reflect the input allelic ratio due to probe cross-hybridization and other confounding factors. Indeed, when all probes were considered, there was no obvious linearity between the measured and the input allele ratios and their values often vary to large extents (Fig. 2A). To screen for blocks that exhibit strong linearity at different input allelic ratio, we plotted the distribution of r values of the 12,950 blocks and observed a skewed normal distribution biased toward large positive values (Fig. 2B). This observation prompted us to perform the t test to examine whether an r value is significantly larger than the standard error, which led to the determination of an r value cutoff at 0.811 ($p = 0.05$; Fig. 2B). Using this cutoff, 519 of the 12,590 blocks were considered to be diagnostic for quantifying the linear effect of input cDNA on the measured allelic ratio (Fig. 2C). Further, the measured ratios from these blocks were also more comparable to the input ratio (comparing Fig. 2C with A). An example of a diagnostic block is illustrated where the measured ratios show an excellent correlation ($r > 0.99$) with the ratio of the template cDNA (Fig. 2D). Importantly, the 519 diagnostic blocks account for only 4.1% of all probe blocks tested yet they represent 284 (55%) of the 518 SNPs (Table 1). Together these results attest to the effectiveness of the tiling array method in uncovering diagnostic blocks for large numbers of SNPs.

2.3. Characterization of the diagnostic blocks

To gain insight into the property of the diagnostic blocks, we first investigated whether they are biased toward certain type of SNPs. In the 518 SNPs, transitions (A:G and C:T) are twice as frequent although there are twice as many possible transversions (A:C, A:T, C:G, and G:T; Fig. 3A). This pattern is highly similar to genome-wide SNP