Chapter 1

A Brief Introduction to Tiling Microarrays: Principles, Concepts, and Applications

Christophe Lemetre and Zhengdong D. Zhang

Abstract

Technological achievements have always contributed to the advancement of biomedical research. It has never been more so than in recent times, when the development and application of innovative cutting-edge technologies have transformed biology into a data-rich quantitative science. This stunning revolution in biology primarily ensued from the emergence of microarrays over two decades ago. The completion of whole-genome sequencing projects and the advance in microarray manufacturing technologies enabled the development of tiling microarrays, which gave unprecedented genomic coverage. Since their first description, several types of application of tiling arrays have emerged, each aiming to tackle a different biological problem. Although numerous algorithms have already been developed to analyze microarray data, new method development is still needed not only for better performance but also for integration of available microarray data sets, which without doubt constitute one of the largest collections of biological data ever generated. In this chapter we first introduce the principles behind the emergence and the development of tiling microarrays, and then discuss with some examples how they are used to investigate different biological problems.

Key words Tiling microarrays, ChIP-chip, Transcriptome mapping, Gene expression, Probe design

1 Introduction

Large-scale analysis of genomic data ushered in a new era of biomedical research. With the ability to investigate whole genomes, it is now possible to evaluate and infer the degree of expression of a large number of genes, or even entire genomes. The latest technological achievements in genomics, biotechnology, informatics, and miniaturization enabled the emergence of the microarray technology. So far, traditional microarrays were mainly used to identify transcripts from samples in a specific context or condition. It then became possible to monitor the level of expression of some particular genes. Knowing the relative abundance of a transcript can indeed reveal the degree of expression of its gene. Their application

Tin-Lap Lee and Alfred Chun Shui Luk (eds.), *Tiling Arrays: Methods and Protocols*, Methods in Molecular Biology, vol. 1067,
DOI 10.1007/978-1-62703-607-8_1,

allowed as well a high-throughput comparison between samples, aiming to identify discrepancy in patterns of expression between different conditions and ultimately identify biomarkers for potential diagnostic or therapeutic purposes.

The microarray technology has its root and founding principles in the Southern blotting technique developed by Edwin Southern in 1975 [1], but it provides a massive scale-up in the nucleic acid hybridization assay because of its huge capacity resulting from miniaturization and automation [2]. The development of cDNA libraries [3] made it possible to spot cDNA probes onto a solid surface inside microscopic wells, allowing the high-throughput screening of thousands of genes simultaneously. In one of the very first microarray studies, Schena et al. in 1995 used spotted DNA microarrays to measure the differential expression of 45 genes in *Arabidopsis thaliana* [3]. Despite such an impressive scale-up, the early DNA microarrays had an obvious technological limitation: because the probes need to be designed to target specific genes [4], the annotation of which may be incomplete and erroneous [2], incorporating annotation data into the experimental design could add errors and bias. As a result, an ideal microarray technology for genome-wide screen should avoid any prior genome or gene annotations.

Tiling microarrays share the same biophysical principle with traditional genomic microarrays by hybridizing labeled fragments of cDNA or RNA to small polymers of nucleotides (probes) attached to solid glass surfaces. However, they use an *unbiased* approach: unlike traditional microarrays, they preclude a priori annotation information. In addition, the probes used in tiling microarrays can overlap, sometimes with only several nucleotide-shifts (e.g., single-base offset) [5, 6]. They also comprehensively represent large portions of genome, or even entire genomes in the case of high-density tiling microarrays (Fig. 1). Instead of probing for known or predicted gene sequences, tiling microarrays map genome sequences that exist contiguously in the genome, or portion of the genome like chromosomes, in a totally unbiased fashion. Today, combined with complete whole-genome sequences, the DNA microarray technology can be used in novel experimental approaches to investigate different biological problems on the whole-genome level [2].

In this chapter, we give a general introduction to tiling microarrays, including their different types and probe designs. We also present their most popular forms of applications with examples and discuss their limitations. The data analysis step is also explored, discussing some of the most common tools employed to analyze the final output from tiling array experiments.

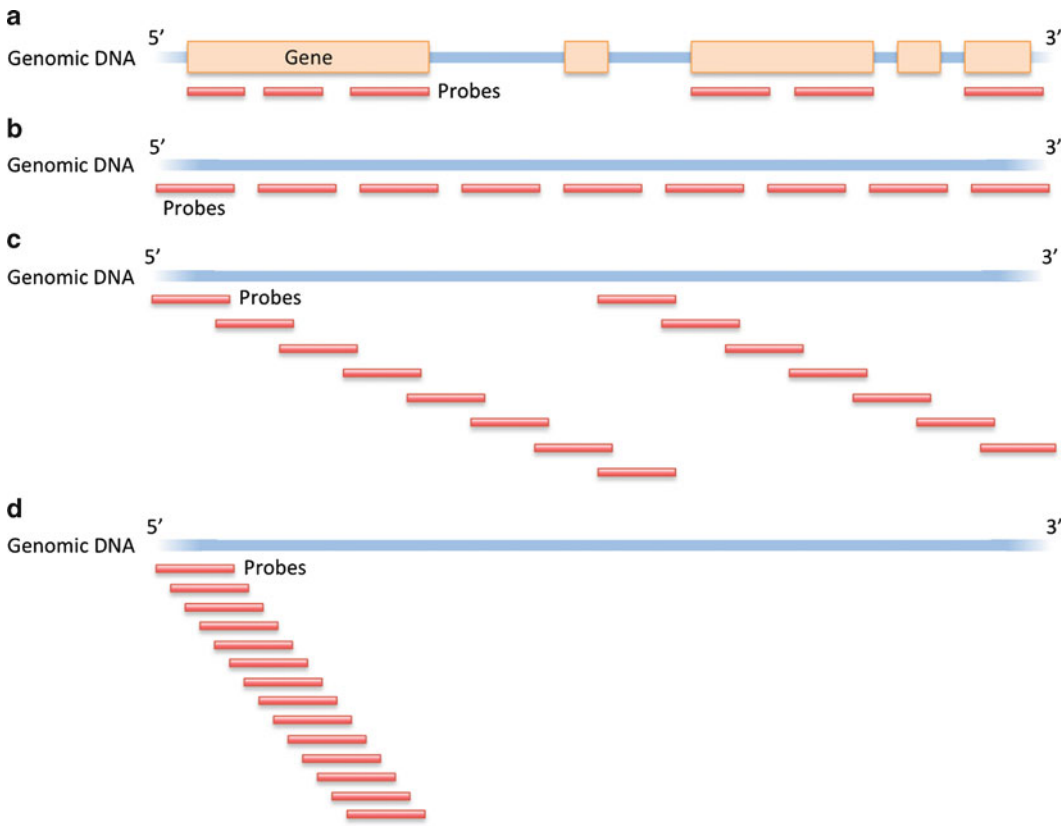

Fig. 1 Different probe designs and tiling paths. (**a**) Classical gene expression microarray. (**b**) Tiling microarray with nonoverlapping probes leaving few base-pair gaps between probes. (**c**) Tiling microarray with overlapping probes. The overlaps are several nucleotides long. (**d**) Tiling microarray with 1-nucleotide overlapping probes, which offers the highest resolution

2 Probe Design and Manufacturing of Tiling Microarrays

Like the original genomic microarrays, tiling arrays rely on the principle of hybridization of labeled molecules of interest to probes attached to a solid surface. This principle, first described in the seminal paper by Watson and Crick in 1953 [7, 8], is the heteroduplex formed by complementary base pairing. However, tiling arrays consist of a more comprehensive and unbiased presentation of the genome than gene microarray. With the constantly increasing number of available complete genome sequences, it is now possible to design contiguous probes without relying on any potentially erroneous annotation [9]. Additionally, the organization of the probes can be arranged for higher or lower resolution depending on the gap and span of overlap between the probes to suit the need of a particular experiment.

2.1 Probe Design

Designing probes to represent a complete or partial genome is the first step in a high-density tiling arrays experiment. It poses a challenge to computational biologists, as they need to consider several aspects of the microarray experiment and answer the following three questions [10–12]:

- How to avoid nonspecific hybridization of molecules to the probes?
- How to maximize hybridization with the biophysical properties of the probe sequences?
- Which probe design is best for the expected outcome of the tiling microarray experiment?

2.1.1 Specificity and Uniqueness of Probes

Due to its nonrandom nature, a genome contains many redundant sequences [10]. Consequently, cross-hybridization between probes and nonspecific DNA/RNA molecules can become a serious problem for a hybridization-based method, as it can result in significant background noise [12]. This problem is especially grave with a genome-wide coverage and becomes even more critical with shorter probe sequences [10]. Alignment-based tools (with BLAST or BLAST-like algorithms [13, 14]) can provide a solution to overcome this problem where potential probes are tested for their specificity by alignment algorithm against a sequence database [10, 15]. RepeatMasker is the most widely used algorithm for repeat detection, identifying repeat regions in genomes using a database of repeat elements [16]. Repetitive sequences can then be masked before probe design.

2.1.2 Thermodynamic and Biophysical Properties of the Probes

An important determinant of how probes behave towards hybridization is their thermodynamic property. This property determines the probe melting temperature and thus, in turn, affects the hybridization process [17]. The probe design also needs to take into consideration the possibility of probe self-hybridization [12]. Another aspect to be considered is the potentially damaging biochemical properties that are required by the manufacturing process [18]. Given the sequence of a particular probe, its thermodynamic characteristics can be determined by well-characterized formulas taking into account either the nucleotide composition of the sequence or the exact sequence in itself [10].

2.1.3 Experimental Design

The final aim of the overall experiment is also critical in the choice of the platform and the design of the probes. A transcriptional mapping study will be more efficient using overlapping tiles and high density [19], whereas copy number variation studies may benefit from a sparse tile path [20]. ChIP-chip methods have also been described to be more efficient with high-density tiling microarrays with consistent spacing of probes, particularly for DNA–protein-binding site identification [21].

Several software solutions have been developed to provide efficient tools for designing high-density tiling array probes with unique sequences. Some solutions have been put forward as integrative platforms in order to optimize the design of probes. For example, Gräf et al. described a method defining a uniqueness score with the content of the shortest substring of the probe and its similarity to any other portion in the genome [12]. The probes designed with this method were subsequently validated with BLAT and further selected with their thermodynamic features taken into account. An alternative presented by Bertone et al. used dynamic programming to infer tile paths and determine most optimal probe sequence using a heuristic approach to reduce dataspace complexity [10]. Recently, ChipD [11] was developed as a Web server-based solution for probe design, providing the user with the sequence of probes, their genomic locations, as well as their hybridization features. However, there is still potential for further improvements in the probe design step of tiling microarray experiments.

2.2 Manufacturing

Depending on their design and manufacturing, tiling microarrays can be classified into two types. The first type of tiling microarrays was manufactured by spotting PCR products as probes on the solid surface, a method also used to make traditional microarrays. This type of tiling microarrays was developed and used in the first half of the 2000s [12]. In some studies, they were successfully used to identify DNA–protein interactions [22, 23] or histone modifications [24] in certain part of the human genome. Despite such successes, the limitations of this type of tiling microarray in genomic coverage and resolution are evident and led to the emergence of a second type. High-density oligonucleotide microarrays consist of short probes (about 100 nucleotides) that are directly synthesized in situ onto the solid surface of the microarray. This synthesis can be performed using a photolithographic mask [25–27], but other techniques exist, such as ink-jet [28]. This type of array can present several millions of spots, each containing several millions of the same probe. This second type has been shown to outperform PCR-product arrays [24].

Several companies, as leaders in the microarray technology, have developed DNA chips for tiling microarray experiments. Currently, Affymetrix has issued a tiling microarray with six million 25-mer probes, offering very high definition. They propose three different platforms for human. The GeneChip Human tiling 1.0 Array Set consists of 14 arrays, designed to cover the entire human genome, devoted for transcript mapping experiment. The GeneChip Human tiling 2.0 Array Set, with 7 arrays, has been designed for ChIP-chip application; and the GeneChipHuman Promoter 1.0 consists of a subset of probes of the previous microarray for promoter regions. These tiling arrays are designed for an optimal specific hybridization, using RepeatMasker [29],

containing over 6.5 million probes per set, and with 25-mer oligonucleotides with about 10-base pair gap between tiles, leading to a 35-base pair resolution [30]. Roche/NimbleGen have also developed their tiling platforms, which usually have lower probe densities than Affymetrix products. For comparative genomic hybridization, they offer nine different platforms with up to 4.2 million 50- to 75-mer probes and 284-bp gaps between probes, offering above 1.4-kbp resolution for CNV discovery. For ChIP-chip experiments, they offer a 10-array set and a 4-array set with 2.1 million probes. They both spot 50- to 75-mer probes with a median gap between probes of 100 bp for the first one and 205 bp for the second. They also offer whole-genome tiling arrays in a similar format for DNA-methylation studies [30]. Similar formats and settings are also available from Agilent, but they usually present lower density. For example, their Human ENCODE ChIP-chip microarray presents over 153,000 probes, and their CGH bundles offer up to 1.1 million probes [30].

3 Applications

3.1 Protein–DNA Interaction Profiling

Gene expression is highly regulated by many mechanisms, particularly via DNA-binding of transcription factors and other molecules. Chromatin immunoprecipitation arrays (ChIP-chip) represent the most popular form of application of tiling microarrays. This particular technique, which couples chromatin immunoprecipitation (ChIP) and microarray techniques (or "chip"), aims to screen on a genome-wide scale the binding sites of transcription factors and other DNA-binding molecules and proteins [31]. Through its determination of transcription factor-binding sites, ChIP-chip can shed new light on regulatory patterns of gene expression [32] and thus provide extremely important information towards a better understanding of critical regulation and gene functions [12].

A ChIP-chip experiment involves cross-linking proteins and their bound DNA by formaldehyde, followed by immune-precipitation, and shearing DNA into ~500-bp fragments (Fig. 2a) [33]. The protein of interest (e.g., a transcription factor), bound to its DNA-binding sites, will precipitate with its bound DNA fragments in the presence of its antibody. The fragments of DNA co-precipitated with the protein of interest are then labeled (e.g., using a fluorescent tag) and hybridized to a microarray. The image is processed to detect the signal of the labeled fragments hybridized on the microarray [34]. The probes with ChIP signals can also be mapped back to the genome [10] in order to identify genome-wide protein-binding sites. Due to the nature of this approach, a major limitation is the quality of the antibody and its affinity to the protein of interest. Indeed, the quality of the result is highly dependent on the quality of the protein precipitation by the antibody [32].

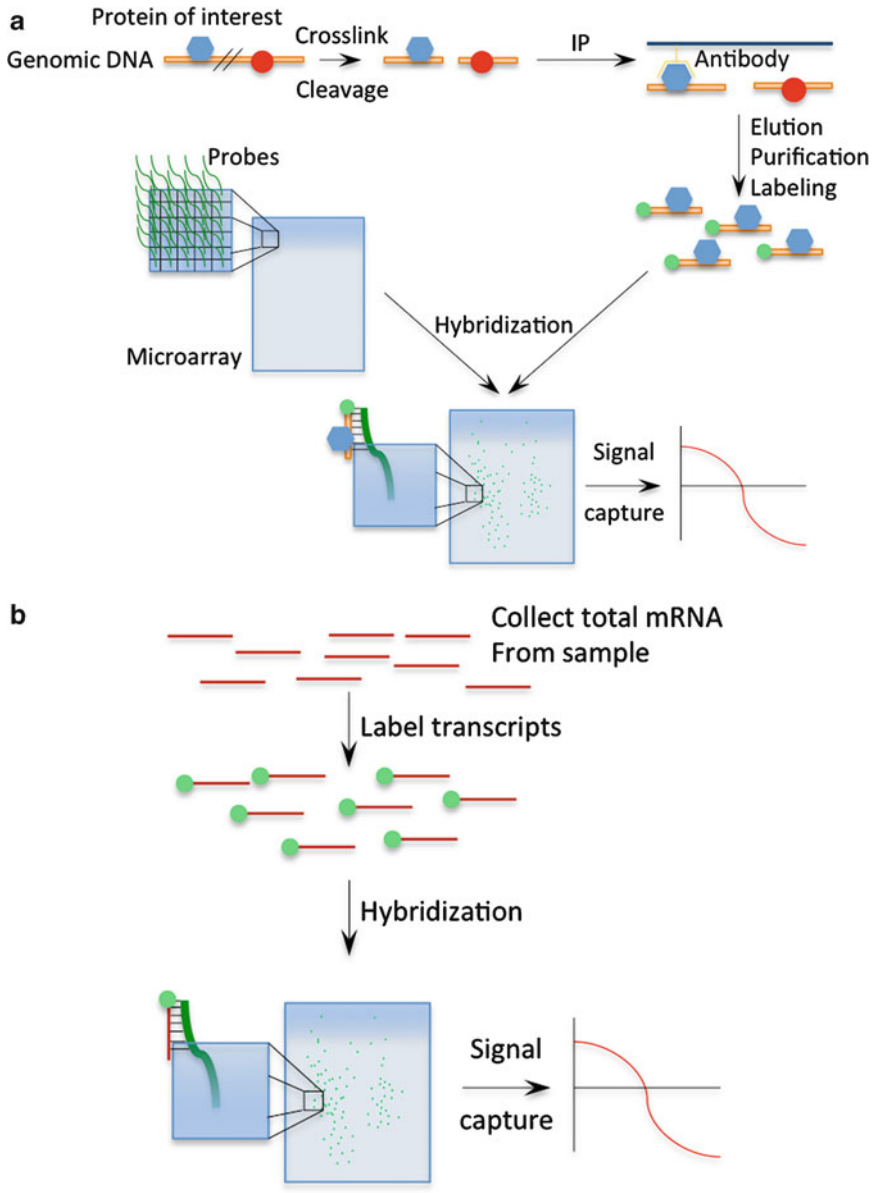

Fig. 2 Schematic workflow summarizing different tiling microarray applications. (**a**) ChIP-chip application of tiling arrays; after cross-link of DNA with proteins, the strands are cleaved and the fragments presented to the antibody specific to the protein of interest (in *blue*). The antibody retains only the fragments with the protein of interest and the others are eluted. The retained fragments linked to the protein are then hybridized to the tiling microarray and mapped to the genome. (**b**) A transcriptome mapping application of tiling microarray where mRNAs from the sample are extracted and labeled and then hybridized to the tiling microarray. (**c**) In nucleosome localization, the internucleosomal DNA is firstly digested with DNaseI, leaving only nucleosomes attached to their DNA. The octamer of protein is then eliminated to only leave out the nucleosomal DNA fragments that are then labeled. The fragments are then hybridized to the microarray and mapped back to the genome. (**d**) An arrayCGH experiment firstly isolates genomic DNA from a patient and from a control, which are then differentially labeled. The two sets of labeled DNA are then presented to the tilling microarray for competitive hybridization. The signal obtained compares the different copy numbers between the two sets of samples and deciphers the copy number variations

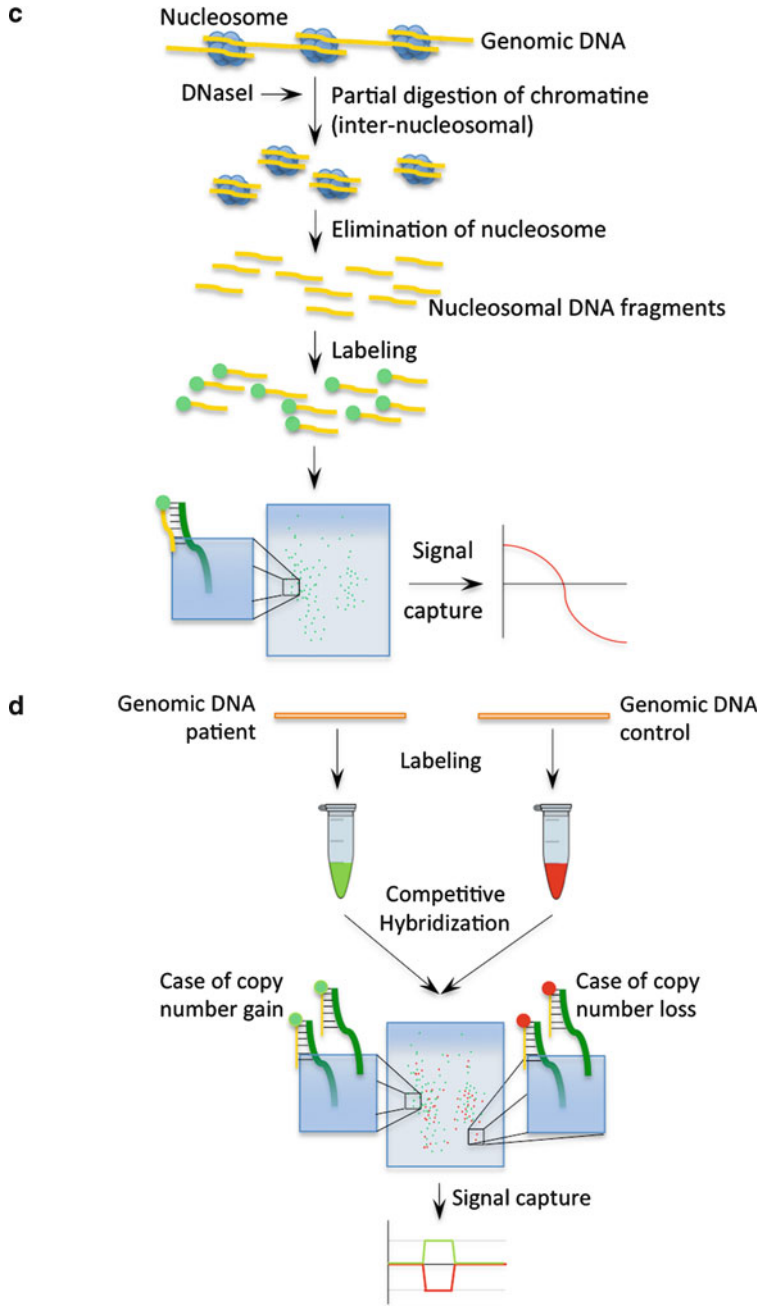

Fig. 2 (continued)

Multiple studies used ChIP-chip approach to identify proteins and transcription factor DNA-binding sites. Amongst the first applications of ChIP-chip were studies screening the genomic binding sites of 16 transcription factors in yeast [35, 36]. In a recent review [37], Peggy Farnham emphasized the great insights

brought by genomic studies using ChIP-chip (and lately ChIP-seq), noting the confirmation that RNA polymerase II and transcription factors bind almost exclusively to proximal promoters that are weakly active, and also that the binding of a transcription factor to enhancer regions may not be sufficient for regulation, hence promoting the idea of a "cooperation" of regulators.

3.2 Gene Expression and Transcriptome Mapping

Gene expression mapping is the other most common application of tiling microarrays. The underlying idea is to hybridize reverse-transcribed RNA into cDNA to tiling microarrays in order to identify the transcribed portions of the genome. Because using tiling microarrays does not involve prior gene annotation, it would be an alternative approach towards gene prediction in a systematic fashion. Traditional in silico gene prediction approaches are known to be error-prone and can miss some rare or condition-specific transcripts [25]. The high resolution of tiling microarrays can overcome some of these limitations. For instance, their sensitivity makes it possible to identify rare transcripts. Also, thanks to their overlapping probes, it is possible to obtain a much more accurate picture of the predicted gene structure.

For this approach, RNAs are extracted, purified (with rRNA depletion), and reverse transcribed into cDNA (Fig. 2b). The cDNA molecules are then labeled and hybridized to a tiling microarray. The transcripts observed can then highlight the portion of the genome from which they originate, therefore providing useful information about the gene position or structure.

As stressed in a review by Mockler & Ecker [25], the power and resolution of this approach for gene and gene structure prediction can be observed with some of the first applications of high-density oligonucleotide tiling arrays to transcription mapping on human chromosomes 21 and 22 [38–40]. They used 25-mer probes with a gap of 35 bp between probes across the entire chromosomes 21 and 22. They were able to demonstrate that it was possible to identify up to ten times more transcripts than what was previously thought and predicted, and that about 90 % of the observed transcription was taking place in non-previously annotated regions. In another early application of tiling microarrays to transcriptome mapping [41], Kapranov et al. identified new classes of small RNAs (sRNAs) using a high-resolution tiling microarray (5-nucleotide resolution). Two of these new types of sRNAs were found to cluster at each extremity of the gene. A third class of sRNAs, which skip most of the exons, was also identified at the 5′ limit of protein-coding genes.

In a recent study, Spencer et al. used tiling microarrays to examine the spatiotemporal motifs of expression of genes in *Caenorhabditis elegans* [42]. Using tiling arrays to profile gene expression in more than 30 cells at different developmental stages of the organism, they discovered novel transcription regions in

noncoding domains that count for about 10 % of the genome. They also showed a differential pattern of expression of about 75 % of the transcripts between the different developmental stages. A similar approach has also been applied to *Drosophila melanogaster* [43].

High-density tiling microarrays can also be used to identify alternatively spliced forms of a transcript. Like DNA-methylation, alternative splicing plays an important role in genetic regulation. It allows a single gene to code multiple functionally distinguished proteins with different amino acid sequences [44]. Moreover, this expansion of the expression potential of a gene is further increased with the different types of alternative splicing that have been observed, including exon skipping or intron retention [32]. Tiling microarrays offer the potential for screening alternative splicing and discover de novo transcript isoforms. In theory, tiling microarrays can identify any type of alternative splicing. However, the probe density can limit the sensitivity of the detection of some alternative splicing events [44]. The unbiased coverage of whole genome offered by tiling microarrays enables the detection of different exons present in the transcripts. Hybridization signals from different exons are useful to elucidate gene structures, and subsequent comparison between signals can help to identify the different isoforms existing in the sample. Recently Eichner et al. presented a novel computational approach based on support vector machines to identify intron retention and exon skipping [44].

3.3 Identification of Genomic Copy Number Variations

Microarray-based comparative genome hybridization (array-CGH) technique uses microarray to compare the genome from the control sample with the disease sample to identify the genomic segments with copy number variations (CNVs). This approach is often chosen for diagnostic purposes, since comparison between two types of samples is possible (e.g., cancer cells versus control cells). The differential hybridization can therefore highlight a potential biomarker of interest that presents a different pattern between the two sets of samples. The methodology is straightforward, consisting of labeling the control and test samples with different fluorescence dyes and hybridizing them to the tiling microarray. The fluorescence is then captured and processed. The difference between fluorescence from the sets of samples is calculated and allows inferring the CNV between the samples (Fig. 2c).

Sebat et al. were amongst the first to report the use of tiling microarrays to identify differences in copy number of genome portions between samples from different tissues and lymphoblastoid cell lines [45]. They observed different patterns between the samples of the same individual and found somatic mutations in gene clusters of T cell receptors and immunoglobulins. In a recent study, Brommesson et al. used array-CGH and hierarchical clustering to show that one out of five pairs of unilateral breast carcinoma shared copy number profile while only one of eight pairs of bilateral

tumors showed similar genomic features [46]. This result shed new light on the genomic origin of unilateral and bilateral primary breast tumors. In their review, Shinawi and Cheung discussed the important applications of array-CGH, stressing the tremendous improvements in diagnosis for numerous conditions including autism and dysmorphism [47]. The possibility to measure the relative difference of abundance of CNVs between two samples has enhanced the potential to discover new patterns and new biomarkers.

3.4 Epigenomics

As a major process of epigenetics, DNA methylation has been intensively studied to establish its involvement in gene expression regulation. With the advent of tiling microarrays, it has been possible to map DNA methylation sites on a genome-wide scale, which is the principle of MeDIP-chip. Similar to ChIP-chip, MeDIP-chip uses immunoprecipitation with an anti-5-methyl-cytosine antibody to specifically precipitate the genomic segments containing methylated DNA and tiling microarray to hybridize the precipitated DNA.

3.4.1 MeDIP-Chip

Studies of DNA methylation with tiling microarrays provided the first so-called *methylomes* and advanced our understanding of how gene expression is affected by epigenetic modifications and regulations. First achieved in *Arabidopsis* [48, 49], these pioneer experiments in methylome discovery showed that highly repetitive DNA sequences presented the highest density of DNA methylation. In a recent study, Rauch et al. used tiling microarrays to examine DNA methylation in human B cells [50] and found that most of the methylated regions were associated with genes.

3.4.2 ChIP-Chip

ChIP-chip experiments can also provide insights into epigenomic components. As demonstrated by the literature, studies of histone modification remain one of the main applications of ChIP-chip. In 2005, Berstein et al. developed a comparative analysis of histone modification between human and mouse using a tiling microarray-based ChIP-chip approach [51].

3.4.3 Nucleosome Localization

Nucleosomes are the basic units in the process of chromatin compaction in eukaryotes, in which DNA strand is wrapped around histone octamers. Tiling microarrays can also be used to study nucleosome localization, which is essential for a better understanding of how DNA-binding proteins such as transcription factors get access to their binding sites on the DNA. This technique involves partial digestion of the chromatin in order to eliminate the linker DNA region between two nucleosomes. After removal of the proteins, the nucleosomal DNA is isolated, labeled, and subsequently hybridized to the tiling microarray (Fig. 2d). This approach requires a high sensitivity and resolution for the tiling microarray due to the size of the portion of nucleosomal DNA (146 or 147 nucleotides) [33].

4 Processing and Analysis of Tiling Microarray Data

The large amount of data generated by a tiling microarray experiment, the complexity, and the framework of the study necessitate applying specific algorithms to extract the relevant information expected from the experiment. Even though many software packages have been developed to analyze complex data sets generated by high-density tiling microarrays, there still remains much room for improvement and innovation in algorithms and methods of tiling microarray data analysis. And indeed, novel approaches still appear in the literature.

Amongst the most popular software tools for tiling microarray data analysis is Tiling Analysis Software (TAS), developed by Affymetrix for its own microarray platforms [52]. With TAS, the differential expression between the probe intensities within a given window is estimated with a Wilcoxon rank test. Kampa et al. also implemented a probe local expression level estimation method in TAS [40]. An alternative to TAS as a peak-detection algorithm is MAT (Model-Based analysis of Tiling arrays) [53]. MAT normalizes the signal from probes using a Mixture Model. It is commonly applied to ChIP-chip data to identify DNA-binding sites, however, other groups have reported their successful adaptation of MAT to transcriptome mapping data [54, 55]. MAT first estimates a baseline of the signal from the probes using a basic linear modeling approach considering two parameters: features from the probe sequences and the copy numbers. This baseline allows pre-filtering most of the noise and reveals the actual significant signal from the probes. Subsequently, MAT standardizes each probe on the microarray using a probe behavior model and the estimate of the baseline, providing a *t*-value representing the actual behavior of the probe from its predicted one and therefore stressing a potential ChIP-enriched portion. TileProbe integrates publicly available data to maximize the outcome from MAT [56]. Besides applications specifically designed for Affymetrix platforms, alternative methods have been developed for NimbleGen microarray platforms. TAMAL has been described as the algorithm of choice for the detection of the binding sites in a ChIP-chip experiment for these platforms [57].

TileScope, an online pipeline for tiling microarray data analysis, is an example of software with a user-friendly interface [58]. It can normalize signals channel-wise and between arrays, calculate a score for each probe, and identify genomic regions with significant signal-to-noise ratios. Starr, another fully integrated pipeline, can be used for exploration, visualization, or quality control [59]. After peak detection in a ChIP-chip experiment, motifs in the transcription factor-binding sites can be identified by MEME [33, 60].

TileShuffle has been recently developed to identify noncoding RNAs in tiling microarray experiments [52]. The algorithm identifies statistically significant, differentially transcribed, and differentially expressed portions from a background distribution. Many statistical algorithms are available for tiling microarray data analysis to identify alternative splicing. MIDAS [61], the most popular method, uses an ANOVA test to distinguish signal from alternatively spliced exons. This statistic test is based on the assumption that the signal intensity of an exon is relatively constant compared to the signal intensity of the overall gene. As mentioned earlier, methods based on machine learning have also been recently developed to analyze tiling microarray data, particularly for alternative splicing detection. For example, Eichner et al. used the support vector machine (SVM), a supervised machine learning technique, to analyze alternative splicing data from tiling microarrays [44].

5 Discussion and Future Directions

Tiling microarrays have been great achievements in biotechnology: for the first time, cells can be queried and analyzed on a whole-genome level and in an systematic fashion. By using no prior annotation data, tiling microarrays offer an full evaluation and screening of protein–DNA interaction, gene expression or structure, and DNA methylation. As a result, they offer a wide panel of potential application from epigenetics to gene expression and structural genetics. They have been successfully applied in these different aspects and lead to important discoveries. Many software tools have been developed for tiling microarray data analysis, each suitable for a specific microarray platform and its application. These software tools facilitate the application of tiling microarrays to address a wide range of different biological questions, whether it is to identify a protein–DNA interaction pattern within the genome or a differential gene expression between samples.

Because tiling microarrays share similar basic principles with traditional genomic microarrays, they also share some of their limitations: both types of microarrays suffer from nonspecific hybridization [52] and narrow signal dynamic range. Intensive normalization is then required to offset both cross-hybridization and background noise [62]. Despite the great promises from the technology of tiling arrays, a literature review of the field does not indicate extensive applications of tiling arrays between 2007 and 2011. This could be explained by the emergence of the next-generation sequencing technologies [63] as shown in Fig. 3. It has also been shown that high-throughput sequencing approaches have clear advantages over tiling array methods [64, 65]. However, tiling

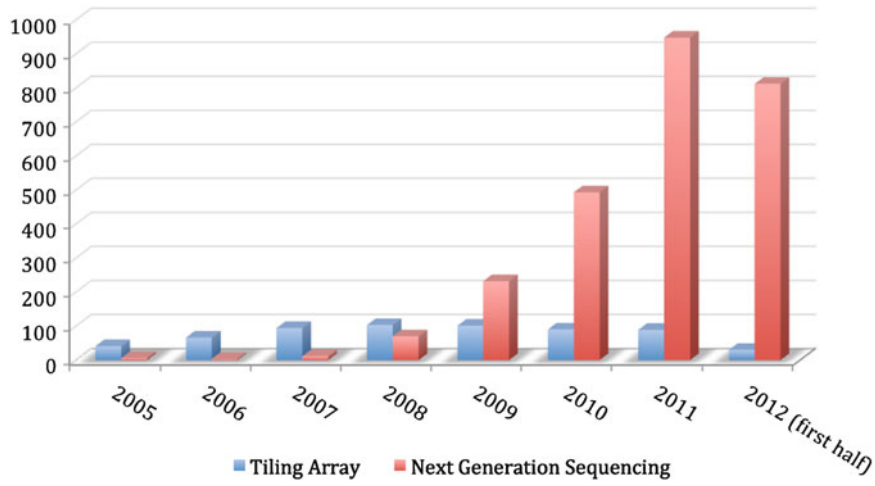

Fig. 3 Number of recent publications related to tiling microarray (*blue bars*) and next-generation sequencing (*red bars*) included PubMed

arrays can still provide an important and meaningful impact in biomedical research, particularly with the development of references and complete genome sequences, and since they have shown to be performing similarly to high-throughput sequencing [63]. In a study in 2009, Sasidharan et al. conducted a comparison between tiling arrays and high-throughput sequencing [63]. This study tried to compare the transcripts identified with tiling arrays and sequencing results, and showed that a significant overlap of identified transcripts exists between the two approaches. But in 2011, Ho et al. published a systemic comparison of a hybridization-based against a sequencing-based ChIP experimental setting (70), in which they showed that ChIP-seq platforms produce better signal-to-noise ratio and better peak-detection with sharper peaks.

From these results and literature, it appears that tiling microarrays have reached a critical step in their evolution, and in the light of alternative innovative approaches, they offer limited advantages. However, they still remain a relatively low-cost option compared to sequencing-based techniques and therefore remain a viable choice for researchers (56), especially when considering they can perform with similar robustness (63).

Acknowledgements

This work was supported by an NIH grant (5R00LM009770-06) from the National Library of Medicine to Z.D.Z.

References

1. Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503–517
2. Lander ES (1999) Array of hope. Nat Genet 21:3–4
3. Schena M et al (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270: 467–470
4. Licatalosi DD, Darnell RB (2010) RNA processing and its regulation: global insights into biological networks. Nat Rev Genet 11: 75–87
5. Kapranov P, Sementchenko VI, Gingeras TR (2003) Beyond expression profiling: next generation uses of high density oligonucleotide arrays. Brief Funct Genomic Proteomic 2:47–56
6. Jacquier A (2009) The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. Nat Rev Genet 10:833–844
7. Watson JD, Crick FH (1953) The structure of DNA. Cold Spring Harb Symp Quant Biol 18:123–131
8. Watson JD, Crick FH (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 171:737–738
9. Yazaki J, Gregory BD, Ecker JR (2007) Mapping the genome landscape using tiling array technology. Curr Opin Plant Biol 10:534–542
10. Bertone P, Trifonov V, Rozowsky JS, Schibert F, Emanuelsson O, Karro J et al (2006) Design optimization methods for genomic DNA tiling arrays. Genome Res 16:271–281
11. Dufour YS, Wesenberg GE, Triltt AJ, Glasner JD, Perna NT, Mitchell JC, Donohue TJ et al (2010) chipD: a web tool to design oligonucleotide probes for high-density tiling arrays. Nucleic Acids Res 38:W321–W325
12. Gräf S, Nielsen FG, Kurtz S, Huynen MA, Birney E, Stunnenberg H, Flicek P (2007) Optimized design and assessment of whole genome tiling arrays. Bioinformatics 23: i195–i204
13. Mount DW (2007) Using the basic local alignment search tool (BLAST). CSH protocols 2007, pdb top17, See http://www.ncbi.nlm.nih.gov/pubmed/21357135
14. Altschul SF, Gish W, Miller W, Myers EW, Lipman DH (1990) Basic local alignment search tool. J Mol Biol 215:403–410
15. Wang X, Seed B (2003) Selection of oligonucleotide probes for protein coding sequences. Bioinformatics 19:796–802
16. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichieme J (2005) Repbase update, a database of eukaryotic repetitive elements. Cytogenet Genome Res 110:462–467
17. Gresham D, Curry B, Ward A, Gordon DB, Brizuela L, Kruglyak L, Botstein D (2010) Optimized detection of sequence variation in heterozygous genomes using DNA microarrays with isothermal-melting probes. Proc Natl Acad Sci USA 107:1482–1487
18. Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS et al (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat Biotechnol 14: 1675–1680
19. Huber W, Toedling J, Steinmetz LM (2006) Transcript mapping with high-density oligonucleotide tiling arrays. Bioinformatics 22:1963–1970
20. Graubert TA, Cahan P, Edwin D, Selzer RR, Richmond TA, Eis PS et al (2007) A high-resolution map of segmental DNA copy number variation in the mouse genome. PLoS Genet 3:e3
21. Qi Y, Rolfe A, Maclsaac KD, Gerber GK, Pokholok D, Zeitlinger J et al (2006) High-resolution computational models of genome binding events. Nat Biotechnol 24:963–970
22. Rada-Iglesias A, Wallerman O, Koch C, Ameur A, Enroth S, Clelland G et al (2005) Binding sites for metabolic disease related transcription factors inferred at base pair resolution by chromatin immunoprecipitation and genomic microarrays. Hum Mol Genet 14:3435–3447
23. Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA et al (2005) A high-resolution map of active promoters in the human genome. Nature 436:876–880
24. Emanuelsson O, Nagalakshmi U, Zheng D, Rozowsky JS, Urban AE, Du J et al (2007) Assessing the performance of different high-density tiling microarray strategies for mapping transcribed regions of the human genome. Genome Res 17:886–897
25. Mockler TC, Chan S, Sundaresan A, Chen H, Jacobsen SE, Ecker JR (2005) Applications of DNA tiling arrays for whole-genome analysis. Genomics 85:1–15
26. Southern EM, Maskos U, Elder JK (1992) Analyzing and comparing nucleic acid sequences by hybridization to arrays of oligonucleotides: evaluation using experimental models. Genomics 13:1008–1017
27. Pease AC, Solas D, Sullivan EJ, Cronin MT, Holmes CP, Fodor SP (1994) Light-generated oligonucleotide arrays for rapid DNA sequence

analysis. Proc Natl Acad Sci USA 91: 5022–5026
28. Hughes TR, Mao M, Jones AR, Burchard J, Marton MJ, Shannon KW et al (2001) Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. Nat Biotechnol 19:342–347
29. Tempel S (2012) Using and understanding RepeatMasker. Methods Mol Biol 859:29–51
30. http://www.home.agilent.com
31. Horak CE, Snyder M (2002) ChIP-chip: a genomic approach for identifying transcription factor binding sites. Methods Enzymol 350: 469–483
32. Gregory BD, Yazaki J, Ecker JR (2008) Utilizing tiling microarrays for whole-genome analysis in plants. Plant J 53:636–644
33. Liu XS (2007) Getting started in tiling microarray analysis. PLoS Comput Biol 3:1842–1844
34. Bernstein BE, Humphrey EL, Liu CL, Schreiber SL (2004) The use of chromatin immunoprecipitation assays in genome-wide analyses of histone modifications. Methods Enzymol 376:349–360
35. Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK et al (2002) Transcriptional regulatory networks in Saccharomyces cerevisiae. Science 298:799–804
36. Xie Z, Hu S, Qian J, Blackshaw S, Zhu H (2011) Systematic characterization of protein-DNA interactions. Cell Mol Life Sci 68: 1657–1668
37. Farnham PJ (2009) Insights from genomic profiling of transcription factors. Nat Rev Genet 10:605–616
38. Cawley S, Bekiranov S, Ng HH, Kapranov P, Sekinger EA, Kampa D et al (2004) Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell 116:499–509
39. Kapranov P, Cawley SE, Drenkow J, Bekiranov S, Strausberg RL, Fodor SP, Gingeras TR (2002) Large-scale transcriptional activity in chromosomes 21 and 22. Science 296: 916–919
40. Kampa D, Cheng J, Kapranov P, Yamanaka M, Brubaker S, Cawley S et al (2004) Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. Genome Res 14:331–342
41. Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT et al (2007) RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science 316:1484–1488
42. Spencer WC, Zeller G, Watson JD, Henz SR, Watkins KL, McWhirter RD et al (2011) A spatial and temporal map of C. elegans gene expression. Genome Res 21:325–341
43. Stolc V, Gauhar Z, Mason C, Halasz G, van Batenburg MF, Rifkin SA et al (2004) A gene expression map for the euchromatic genome of Drosophila melanogaster. Science 306:655–660
44. Eichner J, Zeller G, Laubinger S, Rätsch G (2011) Support vector machines-based identification of alternative splicing in Arabidopsis thaliana from whole-genome tiling arrays. BMC Bioinformatics 12:55
45. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P et al (2004) Large-scale copy number polymorphism in the human genome. Science 305:525–528
46. Brommesson S, Jönsson G, Strand C, Grabau D, Malmström P, Ringnér M, Fernö M, Hedenfalk I (2008) Tiling array-CGH for the assessment of genomic similarities among synchronous unilateral and bilateral invasive breast cancer tumor pairs. BMC Clin Pathol 8:6
47. Shinawi M, Cheung SW (2008) The array CGH and its clinical applications. Drug Discov Today 13:760–770
48. Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H et al (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. Cell 126:1189–1201
49. Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S (2007) Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. Nat Genet 39:61–69
50. Rauch TA, Wu X, Zhong X, Riggs AD, Pfeifer GP (2009) A human B cell methylome at 100-base pair resolution. Proc Natl Acad Sci USA 106:671–678
51. Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ et al (2005) Genomic maps and comparative analysis of histone modifications in human and mouse. Cell 120:169–181
52. Otto C, Reiche K, Hackermuller J (2012) Detection of differentially expressed segments in tiling array data. Bioinformatics 28:1471–1479
53. Johnson WE, Li W, Meyer CA, Gottardo R, Carroll JS, Brown M, Liu XS (2006) Model-based analysis of tiling-arrays for ChIP-chip. Proc Natl Acad Sci USA 103:12457–12462
54. Kadener S, Rodriguez J, Abruzzi KC, Khodor YL, Sugino K, Marr MT 2nd, Nelson S, Rosbash M (2009) Genome-wide identification of targets of the drosha-pasha/DGCR8 complex. RNA 15:537–545
55. Lee SJ, Trostel A, Le P, Harinarayanan R, Fitzgerald PC, Adhya S (2009) Cellular stress

created by intermediary metabolite imbalances. Proc Natl Acad Sci USA 106: 19515–19520

56. Judy JT, Ji H (2009) TileProbe: modeling tiling array probe effects using publicly available data. Bioinformatics 25:2369–2375
57. Bieda M, Xu X, Singer MA, Green R, Farnham PJ (2006) Unbiased location analysis of E2F1-binding sites suggests a widespread role for E2F1 in the human genome. Genome Res 16:595–605
58. Zhang ZD, Rozowsky J, Lam HY, Du J, Snyder M, Gerstein M (2007) Tilescope: online analysis pipeline for high-density tiling microarray data. Genome Biol 8:R81
59. Zacher B, Kuan PF, Tresch A (2010) Starr: simple tiling ARRay analysis of Affymetrix ChIP-chip data. BMC Bioinformatics 11:194
60. Bailey TL, Williams N, Misleh C, Li WW (2006) MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res 34:W369–W373
61. Okoniewski MJ, Yates T, Dibben S, Miller CJ (2007) An annotation infrastructure for the analysis and interpretation of Affymetrix exon array data. Genome Biol 8:R79
62. Sorek R, Cossart P (2010) Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. Nat Rev Genet 11:9–16
63. Sasidharan R, Agarwal A, Rozowsky J, Gerstein M (2009) An approach to comparing tiling array and high throughput sequencing technologies for genomic transcript mapping. BMC Res Notes 2:150
64. Agarwal A, Koppstein D, Rozowsky J, Sboner A, Habegger L, Hillier LW et al (2010) Comparison and calibration of transcriptome data from RNA-Seq and tiling arrays. BMC Genomics 11:383
65. Bradford JR, Hey Y, Yates T, Li Y, Pepper SD, Miller CJ (2010) A comparison of massively parallel nucleotide sequencing with oligonucleotide microarrays for global transcription profiling. BMC Genomics 11:282