Cancer Treatment and Research 158

RNA and Cancer

Bearbeitet von
Jane Y. Wu

1. Auflage 2013. Buch. XIII, 245 S. Hardcover
ISBN 978 3 642 31658 6
Format (B x L): 15,5 x 23,5 cm
Gewicht: 555 g

Detection of Alternatively Spliced or Processed RNAs in Cancer Using Oligonucleotide Microarray

Marieta Gencheva, Lixin Yang, Gong-Biao Lin and Ren-Jang Lin

Abstract

Deregulation of gene expression plays a pivotal role in tumorigenesis, so the ability to detect RNA alterations is of great value in cancer diagnosis and management. DNA microarrays have been used to measure changes in mRNA or microRNA level, but less often the change of RNA isoforms. Here we appraise the utilization of microarray in detecting alternatively processed RNAs, which have alternative splice forms, retained introns, or altered 3′ untranslated regions. We cover the methodology and focus on cancer studies. Recent development in parallel or deep sequencing used in transcriptome analysis is also discussed.

Keywords

Alternative splicing · Splice variants · Intron retention · Alternative 3′ processing · Splicing-sensitive microarray · Genomic tiling microarray · Parallel or deep sequencing

Contents

M. Gencheva · L. Yang · G.-B. Lin · R.-J. Lin (✉)
Department of Molecular Biology, Beckman Research Institute of the City of Hope, 1500 East Duarte Road, Duarte CA 91010-3000, USA
e-mail: rlin@coh.org

J. Y. Wu (ed.), *RNA and Cancer*, Cancer Treatment and Research,
DOI: 10.1007/978-3-642-31659-3_2,

1 Introduction

Cancer is a genetic disease and it is the expression of genetic information, dictated by intrinsic genetic content and by extrinsic influences, that ultimately determines the disease status. Thus, knowing the entire gene activities of cancer cells would have tremendous value to cancer diagnosis and treatment. Gene expression changes are a key feature of cancer development. The changes can be at the level of expression, but they can also be in the forms of RNA due to alternative RNA processing. In this chapter, we illustrate the detection of form changes in RNA due to alternative splicing or 3′ end processing. We include studies that use genomic tiling array to detect unspliced or partially spliced RNA. We also describe parallel or deep sequencing techniques that are recently used to analyze gene expression, and compare them with the microarray approaches.

2 Detection of Alternatively Spliced RNA

2.1 Exon-Junction Microarray

Exon junction microarray is composed of DNA oligos that span the exon–exon junction of a splicing event. Only correctly spliced RNA will stably hybridize to the junction oligo, and thus allowing detection of that splicing event. Exon junction microarray is the first type of array specifically designed to detect alternatively spliced RNA isoforms. The first reported design is used to detect splice events in *S. cerevisiae* and comprises three oligos per gene: one exon probe, one intron probe, and one splice junction probe [1].

To detect exon skipping events in human cells, Shoemaker and colleagues use an array that contains exon junction probes targeting all splice junctions in ~10,000 multi-exon genes [2]. This microarray is used to profile 52 human tissues; they discover evidence of exon skipping in 74 % of those human multi-exon genes, and they also detect a number of previously unidentified splice isoforms. Subsequent microarray designs include exon-body probes, together with exon-junction probes [3–5]. These improved microarrays permit better measurements of RNA isoform quantities. For example, to detect a cassette exon, a probe

set typically consists of exon-body probes to detect the alternative exon and the two flanking exons, as well as probes that monitor each of the three splice junctions. Quantification in the exon-junction arrays depends on the optimal oligo design and probe length, as well as on the data analysis algorithms [4–6]. The design of the junction probe is particularly challenging, since flexibility in choosing the oligo sequences is restricted.

Exon junction microarray is used to determine alternative splicing changes in cell lines derived from Hodgkin lymphoma tumors at different disease stages [7]. Selected genes involved in apoptosis, cell signaling, proliferation, and regulation of splicing are included in the array. Constitutive exons, alternative exons, and splice junctions are monitored by oligonucleotide pairs, one has perfect match and the other contains a single mismatch. Unsupervised clustering of the array data illustrates a clear separation of Hodgkin cell lines and a normal B cell line, with about 20–30 % of the splicing events showing a change in splice isoforms.

In collaboration with the Ares group, we employ exon-junction microarrays to detect alternative splicing changes in two breast cancer cell lines, MCF7 and MDA-MB-231 [8]. We elect to study 64 genes whose alternative splicing is well documented or linked to cancer. For a cassette exon, each alternative splicing event is monitored by oligonucleotide probes, three targeting the splice junctions and three exon-body probes targeting the alternative exon and the two adjacent constitutive exons. For each cassette exon, a skipping index and an inclusion index are calculated, based on the ratio of the signal from junction probes and the signal from constitutive exon probes. A cutoff of 1.5-fold change in the skipping or inclusion indexes is used to detect alternative splicing events. We first compare the two breast cancer cell lines to human mammary epithelial cells (HMEC) and detect a change in splicing pattern in the cancer cell lines for a number of genes, including *hnRNPA/B*, *RBM9*, *FAS*, and *MYL6*. We also identify differentially spliced isoforms between MCF7 and MDA-MB-231 cell lines for *HRMT1L1*, *APLP2*, *CD44*, *VEGF*, *ESR1*, and *EEF1D*. Our microarray study also shows that alternative splicing in tumor xenograft is more closely related to splicing in three dimensional Matrigel cultures than to splicing in two dimensional flat dish cultures.

The use and design improvement of exon junction microarray for detecting alternative splicing have since been described recently including lung cancer studies [9], analysis of the human transcriptome [10], and ABC transporter splicing in drug development [11].

2.2 Ligation-PCR Followed by Microarray Detection

A different method to detect alternatively spliced transcripts is to take advantage of accurate and specific ligation of oligonucleotides hybridizing at the splice junction [12]. The RASL assay (RNA-mediated annealing, selection, and ligation) does not involve prior RNA purification or cDNA synthesis, but is based on detecting a RNA splice junction by annealing to it in solution two oligos complementary to the two respective exonic sequences which create the junction. The correctly

annealed oligos are then ligated and become templates for PCR, which are amplified using universal primers, labeled by either biotin or a dye. The products of the RASL reaction are then hybridized to a fiber optic bead array. This approach has great specificity of detecting a defined set of transcripts by virtually eliminating the problems of cross-hybridization. The method is also more sensitive compared to exon or splice junction arrays that rely on direct hybridization, especially when monitoring small differences between different samples. RASL can detect a well-expressed RNA transcript isolated from less than ten cells, as well as to specifically amplify a transcript from a highly complex RNA mixture.

The RASL method is further improved by first converting mRNA to cDNA in a method called DASL (cDNA-mediated annealing, selection, extension and ligation) [13]. The oligo-annealing and ligation steps are also modified to include a locus-specific oligonucleotide extension. Rather than ligating two oligos at the splice junction as in the case of RASL, this approach permits a gap of 1–20 nt between them, thus providing flexibility for choosing optimal sequences for the oligos and resulted in increasing specificity. The assay is automated and used to detect considerably degraded mRNAs from formalin-fixed and paraffin-embedded tumor samples [14]. The assay is used to profile both transcript abundance and alternatively spliced isoforms in prostate cancer tissues and cell lines [15]. The study identifies a panel of 104 RNA isoforms, the majority of which displays differential expression between normal and tumor prostate tissue and thus can be used as biomarkers. In addition, many alternatively spliced genes linked to prostate tumors are revealed, and differential splicing in *MAPT*, *CACNA1D*, and *AMACR* is validated by RT-PCR. The DASL assay is commercialized by Illumina, which assembles a DASL panel of 502 genes linked to cancer for monitoring gene expression and alternative splicing [16]. The company also supports custom probe panel creation and can profile up to 1,536 targets.

2.3 Exon Microarray

2.3.1 Design and Analysis of Exon Microarray

Since alternatively spliced RNAs have different combinations of exons, one could use quantity changes of individual exons as indicators of alternative splicing. Exon microarrays to detect individual exons in humans and other organisms have been produced. For example, the Affymetrix human exon array contains 5.3 million features (oligo probes) grouped into 1.4 million probe sets. One probe set is typically designed to detect one exon, although some exons have more than one probe set. The array can detect approximately 1.1 million exons or exon clusters, which can be grouped into more than 300,000 different RNAs or transcript clusters. Apart from targeting exons of well known genes, the array also includes probes that map to exons and transcripts supported only by EST or gene prediction algorithms; about half of the probes on the array are based on ESTs or GENESCAN only and offering an opportunity to discover unannotated exons or novel transcripts. The probes are grouped together according to supporting evidence and

the user can choose to look at the signal from only well-annotated exons (core exons, supported by RefSeq and full length GenBank mRNAs), the extended set (ESTs, syntenic rat and mouse mRNAs), or the full set of probes (gene predictions). Exons of less than 25 bases are not represented on the array due to hybridization requirements, so some short exons as well as some alternative 5 or 3′ splice sites are not interrogated by this array.

The exon array can also be used to derive gene level expression data, which is important when analyzing alternative splicing. Since there are no mismatch probes for background estimation, the exon array has to use a different algorithm to estimate nonspecific hybridization by using genomic or antigenomic background probes with defined GC content. Several groups have actually used the human exon array to obtain reliable gene expression data comparable to those from conventional gene arrays [17–20]. Gene expression estimates from the exon array are further improved by selecting only probes with a reliable signal, while removing probes that perform poorly or cross-hybridize to more than one target in the genome [21, 22].

2.3.2 Alternative Splicing in Cancer Detected by Exon Microarray

The exon array has been used to assess alternative splicing in cancer. A study to compare colon tumors with normal tissues has been reported [17]. The authors filter the data both at exon level and at gene level to eliminate outliers and retain exons/transcripts with signal above background. Then they calculate a Splicing Index (SI), which represents the log ratio of exon signals between normal tissue and tumors after normalization to the gene signal. The SI index is used in the MIDAS algorithm (ExACT program, http://www.affymetrix.com/products/software/specific/exact.affx) to test the hypothesis that no alternative splicing occurs for a given exon. This approach identifies a list of potentially differentially spliced genes between normal and tumor colon tissues. Among the 43 genes identified, 9 (*ACTN1*, *VCL*, *CALD1*, *CTTN, TPM1, FN1, COL6A3, SLC3A2,* and *ITGB4)* are validated by RT-PCR. Five of these genes, *ACTN1*, *VCL*, *CALD1*, *CTTN*, and *TPM1*, code for proteins linked to cytoskeleton organization. *FN1* and *COL6A3* are important for the extracellular matrix and *SLC3A2* has a function in integrin signaling. For some of these genes, cancer-specific alternative splice variants have been previously identified. For example *TPM1*, *ACTN1*, and *ITGB4* show similar splicing pattern in colon tumors [23] and *ITGB4* and *TPM1* are identified as differentially spliced in several tumors by computational analysis [24].

A more recent study uses human exon arrays to identify differences in alternative splicing in colon, bladder, and prostate cancer and reports seven genes as differentially spliced: *ACTN1*, *CALD1*, *COLA3*, *LRRFIP2*, *PIK4CB*, *TPM1*, and *VCL* [25]. The validation rate in this study was 67 % (7 out of 15 candidates confirmed by RT-PCR) and the identified genes match the previous studies, including the prevalence of genes involved in the cytoskeletal organization.

Cancer-specific alternative splice forms are identified using exon arrays in tumors of the nervous system. One study [26] reports that *ATP2B4, CaMKII, NLGN4Y, UNC84A, BIN1, MPZL1,* and *NRCAM* are differentially spliced in glial

brain tumors. Glioblastoma samples are used in another study that discovers 14 glioma-specific alternative splicing changes, seven of them novel: *A2BP1*, *BCAS1*, *CACNA1G*, *CLTA*, *KCNC2*, *SNCB*, and *TPD52L2* [27].

The use of exon microarray in detecting alternative splicing has grown in the past few years [28], including studies of breast cancer [29] and method refinements [30].

2.3.3 Improvement in Data Analysis for Alternative Splicing Discovery

The studies based on the Affymetrix exon array analysis tools achieve a relatively low validation rate for detection of differentially spliced exons. Substantial improvements of alternative splicing detection are achieved by two groups using different type of array analysis. One develops a regression-based algorithm for analyzing the array data (REAP), and achieves a validation rate of 60 % for randomly chosen differential splicing events [31]. The other method, MADS (microarray analysis of differential splicing), attempts to remove the major sources of false positives in detecting alternative splicing [32]. The authors incorporate in the analysis algorithm background correction of probes intensities, iterative probe selection for gene expression index calculation, and a procedure to remove sequence-specific cross-hybridization. The method leads to a detection of 25 % more true positive differentially expressed exons compared to Affymetrix's ExACT algorithm, with a validation rate of 90 %. The study also compares the sensitivity of the Affymetrix Exon 1.0 array to a custom spotted oligonucleotide array with splice junction probes [33], and concludes that the latter is more sensitive to small changes in alternative splicing.

3 Detection of Alternatively Processed RNA Using Genomic Tiling Microarray

3.1 The Design of Genomic Tiling Microarray

Genomic tiling arrays aim to interrogate transcripts from a whole genome, using regularly spaced oligonucleotide probes in an unbiased way except for repetitive sequences (reviewed in [34–36]). Three companies offer tiling arrays, with different resolution and hybridization protocols. The Affymetrix protocol uses biotin labeling and one-color assay, while Agilent and NimbleGen produce tiling arrays that use dual-color hybridization and longer oligo probes (60-mer for Agilent, and 50–75-mer for NimbleGen, both tiled at about 100-bp intervals). The Affymetrix tiling arrays are probably the most popular to date and offer the highest resolution. Its human tiling array consists of 25-mer probes with approximately a 10-bp gap between adjacent probes (Fig. 1a). The array has two variants. The GeneChip Human Tiling 1.0 R Array Set includes perfect match (PM) and mismatch (MM) probes in 14 arrays, each having over 6.5 million probes. The GeneChip Tiling 2.0R Array Set has seven arrays, for which the mismatch probes are omitted.

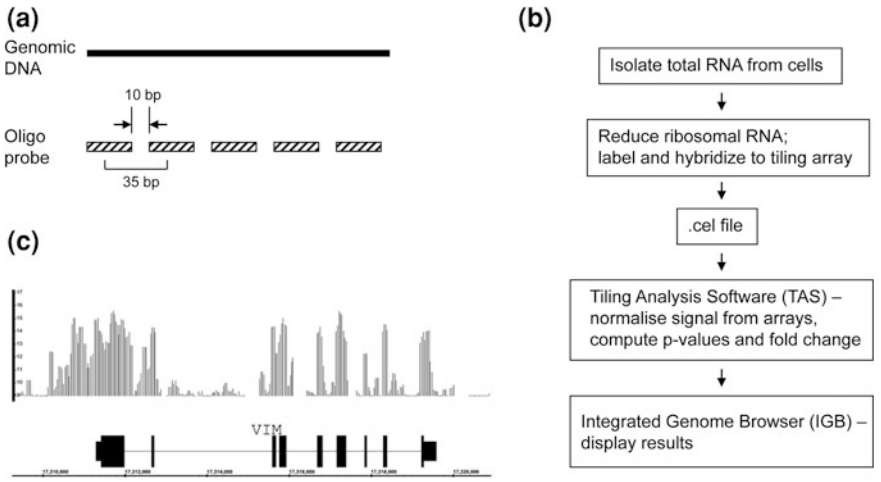

Fig. 1 Design and analysis of a genomic tiling array. (**a**) Probes of 25-mer on a genomic tiling array are spaced at 35 base pairs on the average, with a gap of about 10 bp between adjacent probes. (**b**) Flowchart for tiling array analysis. (**c**) Expression of *VIM* in HEK293 cells transfected with wild type DHX16 cDNA, detected by Affymetrix human tiling array. Exons are represented as boxes and introns as lines. The graph above the gene shows the normalized signal from individual oligo probes as vertical lines, on a log2 scale

Affymetrix has developed Tiling Analysis Software (TAS) for analyzing the arrays, which can perform quantile normalization of the array signal, sliding window analysis of the signal intensities, and subsequent region definition according to user-defined parameters for length and threshold (minimum run and maximum gap). Figure 1b illustrates the steps in a typical experiment involving Affymetrix tiling arrays, and Fig. 1C demonstrates the signal visualization for a single gene, vimentin (*VIM*), taken from an experiment performed in our group (see Sect. 3.3).

3.2 Tiling Microarray Used in the Transcriptome Mapping and the ENCODE Project

The tiling array is unique among the different microarray designs in the fact that it systematically interrogates transcription outside known or predicted genes. Initial experiments to assess transcription across human chromosomes 21 and 22 using tiling arrays are performed with 11 tumor tissues and fetal cell lines and reveal surprising complexity [37, 38]. About an order of magnitude more of the genomic sequence is detected as transcripts than what has been assumed on the basis of annotated and predicted exons. This finding is confirmed and extended by using a tiling array of 36-mer probes at a resolution of 46 base pairs on the average, covering the entire human genome [39]. The study detects numerous transcriptionally active regions (TARs) in addition to unannotated genes, antisense transcription, and previously undetected exons of known genes. A study using arrays

with a high resolution of five base pairs further provides evidence of an extensive network of overlapping transcripts with almost half of all transcribed sequences not polyadenylated [40]. The massive unannotated transcription is not unique to human cells, since similar findings are documented in mouse [41], fly [42], and *Arabidopsis* [43].

Tiling arrays are also used in the ENCODE project [44] to assess transcription of ~1 % of the human genome. The ENCODE tiling array has ~750,000 25-mer PM and MM oligo probes, spaced at 21-bp intervals. The ENCODE study detects transcription fragments (TxFrags) from 14.7 % of the nucleotides represented on the arrays, with 63 % of the TxFrags residing in intronic or intergenic regions. By a combination of 5′RACE and tiling array hybridization, more than 60 % of the annotated protein-coding genes show evidence of new alternative exons in their introns, and 68 % exhibit a potential new transcription start site upstream of their first annotated exon [45]. The distal 5′ exons often overlap with adjacent genes, thus creating chimeric transcripts. Many of the novel exons are expressed in a tissue-specific manner, which underlines the importance of whole-genome transcriptome studies in disease models.

3.3 Detection of Unspliced or Partially Spliced RNA by Tiling Microarray

While alternative cassette exons and many 5 or 3′ alternative splice sites can be detected by exon or exon-junction arrays, unspliced or partially spliced transcripts are best followed by probes that cover the intron regions. The genomic tiling arrays thus provide an opportunity to systematically assess intron retention events arising from perturbation of the splicing machinery. Such types of splicing changes may be important for cancer, since miss-splicing is linked to cancer [46, 47]. While alternative splicing involving exon skipping is underrepresented in tumor cells than in normal cells, intron retention events are at a higher level [48].

Detection of introns by tiling arrays is reported by comparing wild-type *S. cerevisiae* and a mutant strain deficient in processing of excised introns [49]. The study confirms previously predicted introns and discovers new intron-containing genes. A recent study also uses tiling arrays to investigate the effect of nonsense-mediated decay and nuclear exosome on the intron content in *S. cerevisiae* [50]. The study shows that about a third of the yeast introns increase upon inactivation of the nonsense-mediated decay pathway, an effect which is not observed when analyzing yeast intron content by other types of microarrays. The sensitivity of the tiling array detection in this experiment is comparable to Northern blot data.

Our group has used the Affymetrix human tiling array to detect transcripts affected by a human DExH-box spliceosomal protein, DHX16 [51]. RNA samples from cells expressing wild type DHX16 and from cells expressing a dominant negative mutant were removed of ribosomal RNA, labeled, and hybridized to probes on the array. The data were analyzed using TAS to integrate neighboring oligo signals and to identify RNA fragments or intervals. In this analysis, we included only

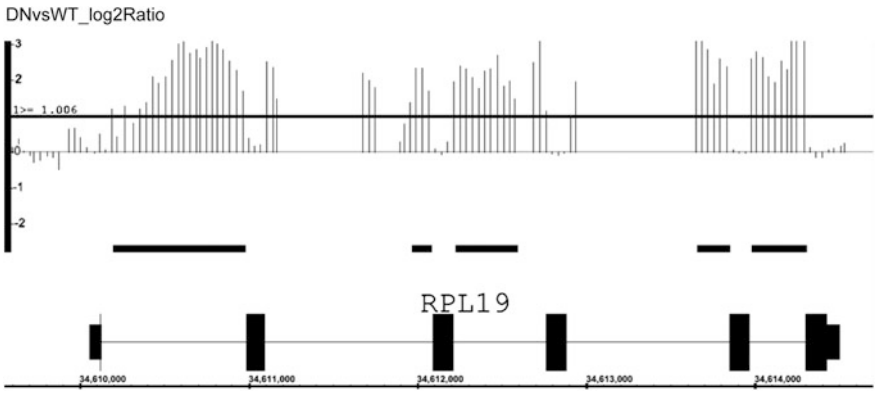

Fig. 2 Comparing RNA from mutant and wild type DHX16-expressing cells using genomic tiling microarray. HEK293 cells were transfected with a dominant negative DHX16 mutant cDNA (DN) or with a wild type DHX16 cDNA (WT). RNA was extracted and analyzed on tiling microarray. Shown here is the genomic region containing *RPL19* gene, which is transcribed from left to right. Log2 of the signal ratio between DN and WT are shown as vertical lines above the gene. Signals that are lower in DN have lines pointing downward. Intervals with a 2-fold increase are depicted with thick bars

intervals with a minimum length of 100 nucleotides, since an average human intron is longer than 4,000 nt and few introns are shorter than 100 nt [52]. We detected a number of genes showing clear evidence of intron retention in the transcripts. *RPL19* is an example of a gene whose transcripts retain introns in the mutant-expressing cells (Fig. 2). The array data indicate that many of those genes have elevated signal from the majority of their introns, a pattern which is possible to detect with the unbiased tiling array. Thus, genomic tiling microarray is effective in identifying gene transcripts that retain introns when splicing is impaired.

3.4 Detection of Alternatively Processed RNAs in Breast Cancer Using Tiling Array

We also used the aforementioned Affymetrix genomic tiling microarray to compare RNA samples from breast tumors and from normal tissues [53]. Breast tumors were biopsies from patients and normal breast tissues were taken from plastic surgery of normal individuals for cosmetic purpose. In this analysis, we included intervals with a minimum length of 50 bases, since the median of human exons is only 124 bases [54]. Intervals representing increase in breast tumors by more than 2-fold with a p value less than 0.01 were further analyzed.

Most of the up-regulated intervals were coincided with annotated mRNA exons. For example, *KPNA2* has 11 exons; five of the exons were identified as up-regulated intervals (Fig. 3a). The remaining six exons of *KPNA2* had probes showing significant increase in signals. KPNA2, karyopherin alpha2, is a potential prognostic marker in breast tumors and predicts poor survival in breast cancer patients [55, 56].

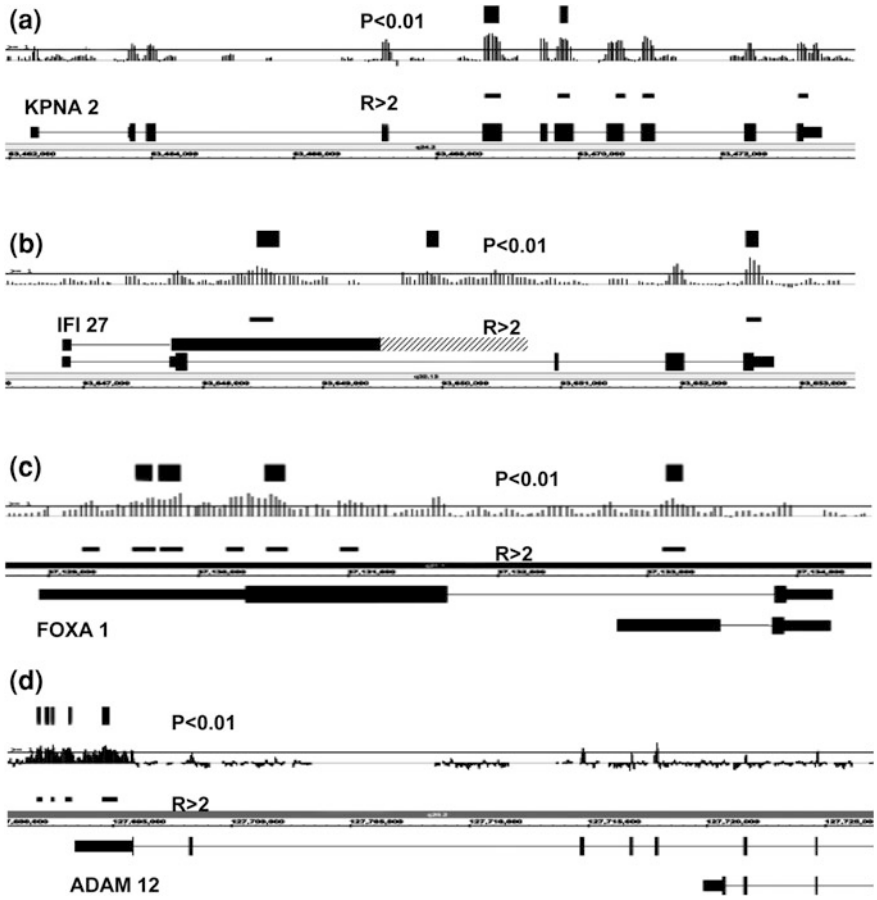

Fig. 3 Comparing the tiling microarray signals between breast tumors and the normal tissues. Four genes are depicted: (**a**) *KPNA2*, (**b**) *IFI27*, (**c**) *FOXA1*, and (**d**) *ADAM12*. The exons and introns are depicted as in Fig. 1; KPNA2 and IFI27 are transcribed from *left* to *right*, while the other two genes are *right* to *left*. The vertical lines above the gene represent log2 of the signal ratio between tumors and normal tissues, with a horizontal line representing 2-fold up-regulation. Intervals with a ratio greater than 2 ($R > 2$) are shown as bars below the ratio lines and intervals with a p value less than 0.01 ($p < 0.01$) are shown above the ratio lines

A few upregulated intervals fell into introns of known genes. For example, an intronic interval was found in intron 2 of the *IFI27* gene. Further analysis indicates that this intragenic interval is part of a novel transcript with a long exon 2 that ends within the intron 2 of the standard, longer transcript (Fig. 3b). Another interval was found in the intron of *FOXA1* gene, which is shown to be an alternative exon (Fig. 3c). IFI27, interferon alpha-inducible protein 27, is up-regulated in a number of epithelial cancers, although its association with breast cancer is less clear [57, 58]. FOXA1, forkhead-box A1, is expressed in breast cancer, although its value in prognostic prediction is still being investigated [59, 60].

Several intergenic intervals were found to locate near the 3′ end of a protein-coding gene. These RNA segments could represent a 3′ extension of the nearby gene or a complete separate transcript. For example, RefSeq database shows transcripts of the *ADAM12* gene having two different 3′ ends (Fig. 3d). The tiling array data indicated that the short form was not well expressed in breast tumor. Moreover, the intervals found 3′ to the *ADAM12* gene, upon further analysis, represent a transcript that is longer than the long form. ADAM12, a disintegrin and metalloprotease 12, is associated with several cancers including breast [61, 62], however, this "extra" long transcript has never been described.

Recently, genomic tiling microarray was used to detect aberrant processing of RNA transcripts at cryptic polyadenylation sites in introns when U1 snRNP was knockdown [63].

4 Transcriptome Analysis by Direct Sequencing

Whole genome direct sequencing of transcripts has emerged as a powerful alternative to microarray analysis [64–69]. It is based on a new generation of massive parallel sequencing technologies. Currently three deep sequencing platforms have gained popularity. The first platform was introduced by 454 Life Sciences (available trough Roche) and relies on pyrosequencing by synthesis. The 454 platform can read 200–400 bases from each molecule and can achieve 1 million reads per run. Illumina offers a bead-based sequencing by synthesis employing reversible fluorescent terminators. The technology was developed by Solexa and can achieve $\sim$3 billion bases per run, with read length of 35–70 bases. A relatively new system is SOLiD (Applied Biosystems), based on massive, parallel sequential ligation technology. SOLiD can map 4–6 billion bases per run, with a read length of 50 bases. Although the length of the reads in all three cases is short compared to the Sanger sequencing, the new technologies provide very high accuracy and ultra-high throughput, making the sequencing of whole transcriptome fast and increasingly affordable.

Analysis of transcripts by sequencing has important advantages compared to microarrays. First, sequencing can discover new isoforms without prior knowledge of the exact sequence. Second, sequencing bypasses hybridization and thus eliminates problems associated with background and cross-hybridization, a major cause of signal variability in microarray analysis [70, 71]. Third, transcript detection achieved by deep sequencing and RNA-Seq protocol is shown to be quantitative with a linear range over five orders of magnitude [65, 66]. Furthermore, it is reported that about 40 million reads of $\sim$25 bases are sufficient to accurately detect splice isoforms for transcripts with high- or moderate expression [65]. Detecting alternatively spliced RNAs, for nearly all of the multi-exon RefSeq human genes with an accuracy of detection comparable to quantitative PCR have also been reported [66].

Massive parallel sequencing has been used to detect cancer-specific transcripts in two comprehensive studies of pancreatic cancer and glioblastoma [72, 73]. In both studies, deep sequencing on an Illumina/Solexa platform is combined with SAGE to

quantitate gene expression. This approach identifies 541 genes differentially expressed in prostate cancers compared to normal cells. On average, these genes have 88-fold higher expression in primary pancreatic tumors. In the case of glioblastoma, 143 genes are determined to be expressed at 10-times the level in normal brain. Many of these genes encoded proteins which are secreted or expressed on the cell surface, thus making them potential diagnostic markers or therapy targets.

5 Summary and Conclusion

In this chapter, we evaluate microarrays that are used to detect alternatively spliced or processed RNA. Ligation of junction-hybridizing oligos followed by PCR and array detection is most sensitive, although the number of splicing events it can simultaneously measure is somewhat limited. Exon junction microarray can measure splice variants with a potential to be genome-wide; however, its accuracy can be affected by the hybridization behaviors of the restricted exon junction probes. All-exon microarray has already reached genome-wide scale and is commercially available; however, there is still a need to develop an algorithm for analyzing most if not all alternatively splicing events reliably.

No microarray is specifically designed to discern RNA with intron retention or alternative 3′ processing in human cells. We describe here our studies using genomic tiling microarray from Affymetrix for these purposes. Although novel alternatively spliced RNA is detected in our studies, the recognition relies on visual examination of the intervals on Integrated Genome Browser and additional molecular assays. The suitability of using genomic tiling microarray for alternative splicing studies remains to be further evaluated. Parallel or deep sequencing has shown great promise in identifying novel transcripts and obtaining quantitative information. However, the storage and computation power needed to handle the massive data generated from sequencing still post a major challenge to its general applications.

The role of alternative or aberrant splicing in cancer is quite evident [74], but more investigations are warranted [75]. The 3′ untranslated region is critical for the translation, stability, localization of the RNA, and microRNA regulation [76, 77], and the length regulation of 3′ UTR in cancer has just gained momentum [78]. Thus, detecting and studying alteration in splicing/processing shall remain a focus in cancer biology.

References

1. Clark TA, Sugnet CW, Ares M Jr (2002) Genome-wide analysis of mRNA processing in yeast using splicing-specific microarrays. Science 296:907–910
2. Johnson JM, Castle J, Garrett-Engele P et al (2003) Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. Science 302:2141–2144
3. Pan Q, Shai O, Misquitta C et al (2004) Revealing global regulatory features of mammalian alternative splicing using a quantitative microarray platform. Mol Cell 16:929–941

4. Fehlbaum P, Guihal C, Bracco L et al (2005) A microarray configuration to quantify expression levels and relative abundance of splice variants. Nucleic Acids Res 33:e47
5. Srinivasan K, Shiue L, Hayes JD et al (2005) Detection and measurement of alternative splicing using splicing-sensitive microarrays. Methods 37:345–359
6. Castle J, Garrett-Engele P, Armour CD et al (2003) Optimization of oligonucleotide arrays and RNA amplification protocols for analysis of transcript structure and alternative splicing. Genome Biol 4:R66
7. Relogio A, Ben-Dov C, Baum M et al (2005) Alternative splicing microarrays reveal functional expression of neuron-specific regulators in Hodgkin lymphoma cells. J Biol Chem 280:4779–4784
8. Li C, Kato M, Shiue L et al (2006) Cell type and culture condition-dependent alternative splicing in human breast cancer cells revealed by splicing-sensitive microarrays. Cancer Res 66:1990–1999
9. Pio R, Blanco D, Pajares MJ et al (2010) Development of a novel splice array platform and its application in the identification of alternative splice variants in lung cancer. BMC genomics 11:352
10. Zhou W, Calciano MA, Jordan H et al (2009) High resolution analysis of the human transcriptome: detection of extensive alternative splicing independent of transcriptional activity. BMC Genet 10:63
11. Srinivasan S, Bingham JL, Johnson D (2009) The ABCs of human alternative splicing: a review of ATP-binding cassette transporter splicing. Curr Opin Drug Discov Devel 12:149–158
12. Yeakley JM, Fan JB, Doucet D et al (2002) Profiling alternative splicing on fiber-optic arrays. Nat Biotechnol 20:353–358
13. Fan JB, Yeakley JM, Bibikova M et al (2004) A versatile assay for high-throughput gene expression profiling on universal array matrices. Genome Res 14:878–885
14. Zhang C, Li HR, Fan JB et al (2006) Profiling alternatively spliced mRNA isoforms for prostate cancer classification. BMC Bioinformatics 7:202
15. Li HR, Wang-Rodriguez J, Nair TM et al (2006) Two-dimensional transcriptome profiling: identification of messenger RNA isoform signatures in prostate cancer from archived paraffin-embedded cancer specimens. Cancer Res 66:4079–4088
16. Fan JB, Gunderson KL, Bibikova M et al (2006) Illumina universal bead arrays. Methods Enzymol 410:57–73
17. Gardina PJ, Clark TA, Shimada B et al (2006) Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. BMC Genomics 7:325
18. Robinson MD, Speed TP (2007) A comparison of Affymetrix gene expression arrays. BMC Bioinformatics 8:449
19. Abdueva D, Wing MR, Schaub B, Triche TJ (2007) Experimental comparison and evaluation of the Affymetrix exon and U133Plus2 Gene Chip arrays. PLoS ONE 2:e913
20. Okoniewski MJ, Hey Y, Pepper SD, Miller CJ (2007) High correspondence between Affymetrix exon and standard expression arrays. Biotechniques 42:181–185
21. Xing Y, Kapur K, Wong WH (2006) Probe selection and expression index computation of Affymetrix Exon Arrays. PLoS ONE 1:e88
22. Kapur K, Xing Y, Ouyang Z, Wong WH (2007) Exon arrays provide accurate assessments of gene expression. Genome Biol 8:R82
23. Okumura M, Kondo S, Ogata M et al (2005) Candidates for tumor-specific alternative splicing. Biochem Biophys Res Commun 334:23–29
24. Xu Q, Lee C (2003) Discovery of novel splice forms and functional analysis of cancer-specific alternative splicing in human expressed sequences. Nucleic Acids Res 31:5635–5643
25. Thorsen K, Sorensen KD, Brems-Eskildsen AS et al (2008) Alternative splicing in colon, bladder, and prostate cancer identified by exon array analysis. Mol Cell Proteomics 7:1214–1224

26. French PJ, Peeters J, Horsman S et al (2007) Identification of differentially regulated splice variants and novel exons in glial brain tumors using exon expression arrays. Cancer Res 67:5635–5642
27. Cheung HC, Baggerly KA, Tsavachidis S et al (2008) Global analysis of aberrant pre-mRNA splicing in glioblastoma using exon expression arrays. BMC Genomics 9:216
28. Hallegger M, Llorian M, Smith CW (2010) Alternative splicing: global insights. The FEBS J 277:856–866
29. Lapuk A, Marr H, Jakkula L et al (2010) Exon-level microarray analyses identify alternative splicing programs in breast cancer. Mol Cancer Res 8:961–974
30. Subbaram S, Kuentzel M, Frank D et al (2010) Determination of alternate splicing events using the Affymetrix Exon 1.0 ST arrays. Method Mol Biol 632:63–72
31. Yeo GW, Xu X, Liang TY et al (2007) Alternative splicing events identified in human embryonic stem cells and neural progenitors. PLoS Comput Biol 3:1951–1967
32. Xing Y, Stoilov P, Kapur K et al (2008) MADS: a new and improved method for analysis of differential alternative splicing by exon-tiling microarrays. RNA 14:1470–1479
33. Boutz PL, Stoilov P, Li Q et al (2007) A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. Genes Dev 21:1636–1652
34. Mockler TC, Chan S, Sundaresan A et al (2005) Applications of DNA tiling arrays for whole-genome analysis. Genomics 85:1–15
35. Johnson JM, Edwards S, Shoemaker D, Schadt EE (2005) Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments. Trends Genet 21:93–102
36. Yazaki J, Gregory BD, Ecker JR (2007) Mapping the genome landscape using tiling array technology. Curr Opin Plant Biol 10:534–542
37. Kapranov P, Cawley SE, Drenkow J et al (2002) Large-scale transcriptional activity in chromosomes 21 and 22. Science 296:916–919
38. Kampa D, Cheng J, Kapranov P 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
39. Bertone P, Stolc V, Royce TE et al (2004) Global identification of human transcribed sequences with genome tiling arrays. Science 306:2242–2246
40. Cheng J, Kapranov P, Drenkow J et al (2005) Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. Science 308:1149–1154
41. Okazaki Y, Furuno M, Kasukawa T et al (2002) Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature 420:563–573
42. Stolc V, Gauhar Z, Mason C et al (2004) A gene expression map for the euchromatic genome of *Drosophila melanogaster*. Science 306:655–660
43. Yamada K, Lim J, Dale JM et al (2003) Empirical analysis of transcriptional activity in the *Arabidopsis* genome. Science 302:842–846
44. Birney E, Stamatoyannopoulos JA, Dutta A et al (2007) Identification and analysis of functional elements in 1 % of the human genome by the ENCODE pilot project. Nature 447:799–816
45. Denoeud F, Kapranov P, Ucla C et al (2007) Prominent use of distal 5' transcription start sites and discovery of a large number of additional exons in ENCODE regions. Genome Res 17:746–759
46. Le Hir H, Charlet-Berguerand N, de Franciscis V, Thermes C (2002) 5'-End RET splicing: absence of variants in normal tissues and intron retention in pheochromocytomas. Oncology 63:84–91
47. Lee MP, Feinberg AP (1997) Aberrant splicing but not mutations of TSG101 in human breast cancer. Cancer Res 57:3131–3134
48. Kim E, Goren A, Ast G (2008) Insights into the connection between cancer and alternative splicing. Trends Genet 24:7–10

49. Juneau K, Palm C, Miranda M, Davis RW (2007) High-density yeast-tiling array reveals previously undiscovered introns and extensive regulation of meiotic splicing. Proc Natl Acad Sci USA 104:1522–1527
50. Sayani S, Janis M, Lee CY, Toesca I, Chanfreau GF (2008) Widespread impact of nonsense-mediated mRNA decay on the yeast intronome. Mol Cell 31:360–370
51. Gencheva M, Lin TY, Wu X et al (2010) Nuclear retention of unspliced pre-mRNAs by mutant DHX16/hPRP2, a spliceosomal DEAH-box protein. J Biol Chem 285:35624–35632
52. Sakharkar MK, Perumal BS, Sakharkar KR, Kangueane P (2005) An analysis on gene architecture in human and mouse genomes. In Silico Biol 5:347–365
53. Yang L, Lin G, Wu X, Yen Y, Lin R-J Novel RNAs overexpressed in breast tumors as revealed by genomic tiling microarray. (In Preparation)
54. Gudlaugsdottir S, Boswell DR, Wood GR, Ma J (2007) Exon size distribution and the origin of introns. Genetica 131:299–306
55. Dahl E, Kristiansen G, Gottlob K et al (2006) Molecular profiling of laser-microdissected matched tumor and normal breast tissue identifies karyopherin alpha2 as a potential novel prognostic marker in breast cancer. Clin Cancer Res 12:3950–3960
56. Gluz O, Wild P, Meiler R et al (2008) Nuclear karyopherin alpha2 expression predicts poor survival in patients with advanced breast cancer irrespective of treatment intensity. Int J Cancer 123:1433–1438
57. Suomela S, Cao L, Bowcock A, Saarialho-Kere U (2004) Interferon alpha-inducible protein 27 (IFI27) is upregulated in psoriatic skin and certain epithelial cancers. J Invest Dermatol 122:717–721
58. Wenzel J, Tomiuk S, Zahn S et al (2008) Transcriptional profiling identifies an interferon-associated host immune response in invasive squamous cell carcinoma of the skin. Int J Cancer 123:2605–2615
59. Badve S, Turbin D, Thorat MA, Morimiya A, Nielsen TO, Perou CM, Dunn S, Huntsman DG, Nakshatri H (2007) FOXA1 expression in breast cancer-correlation with luminal subtype A and survival. Clin Cancer Res 13:4415–4421
60. Habashy HO, Powe DG, Rakha EA, Ball G, Paish C, Gee J, Nicholson RI, Ellis IO (2008) Forkhead-box A1 (FOXA1) expression in breast cancer and its prognostic significance. Eur J Cancer 44:1541–1551
61. Mochizuki S, Okada Y (2007) ADAMs in cancer cell proliferation and progression. Cancer Sci 98:621–628
62. Rocks N, Paulissen G, El Hour M et al (2008) Emerging roles of ADAM and ADAMTS metalloproteinases in cancer. Biochimie 90:369–379
63. Kaida D, Berg MG, Younis I, Kasim M, Singh LN, Wan L, Dreyfuss G (2010) U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. Nature 468:664–668
64. Sultan M, Schulz MH, Richard H et al (2008) A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science 321:956–960
65. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5:621–628
66. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB (2008) Alternative isoform regulation in human tissue transcriptomes. Nature 456:470–476
67. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR (2008) Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. Cell 133:523–536
68. Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, Penkett CJ, Rogers J, Bahler J (2008) Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature 453:1239–1243
69. Cloonan N, Forrest AR, Kolle G et al (2008) Stem cell transcriptome profiling via massive-scale mRNA sequencing. Nat Methods 5:613–619

70. Okoniewski MJ, Miller CJ (2006) Hybridization interactions between probe sets in short oligo microarrays lead to spurious correlations. BMC Bioinformatics 7:276
71. Hoen PAt, Ariyurek Y, Thygesen HH, et al (2008) Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. Nucleic Acids Res 36(21):e141
72. Parsons DW, Jones S, Zhang X et al (2008) An integrated genomic analysis of human glioblastoma multiforme. Science 321:1807–1812
73. Jones S, Zhang X, Parsons DW, Lin JC et al (2008) Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science 321:1801–1806
74. Venables JP (2006) Unbalanced alternative splicing and its significance in cancer. Bioessays 28:378–386
75. Pajares MJ, Ezponda T, Catena R, Calvo A, Pio R, Montuenga LM (2007) Alternative splicing: an emerging topic in molecular and clinical oncology. Lancet Oncol 8:349–357
76. Moore MJ (2005) From birth to death: the complex lives of eukaryotic mRNAs. Science 309:1514–1518
77. Jan CH, Friedman RC, Ruby JG, Bartel DP (2011) Formation regulation and evolution of Caenorhabditis elegans 3'UTRs. Nature 469:97–101
78. Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB (2008) Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. Science 320:1643–1647