

# Chapter 3

## Whole-Genome Microarrays: Applications and Technical Issues

Brian D. Gregory and Dmitry A. Belostotsky

### Abstract

DNA microarrays have become a mainstream tool in experimental plant biology. The constant improvements in the technological platforms have enabled the development of the tiling DNA microarrays that cover the whole genome, which in turn catalyzed the wide variety of creative applications of such microarrays in the areas as diverse as global studies of genetic variation, DNA-binding proteins, DNA methylation, and chromatin and transcriptome dynamics. This chapter attempts to summarize such applications as well as discusses some technical and strategic issues that are particular to the use of tiling microarrays.

**Key words:** Tiling arrays, transcriptome, tilemap, *Arabidopsis*, rice.

---

### 1. Overview of DNA Microarray Technology

The ever-increasing abundance of available genome sequences has enabled a wide variety of experimental and/or computational studies at the whole-genome level. In parallel with the advances in available sequence data, recent improvements in microarray technologies have made it feasible to interrogate a complete genome sequence with arrays through the use of high-density whole-genome tiling microarrays. These DNA microarrays serve as a powerful platform for numerous experimental approaches aiming to probe, in a single experiment, the depths of functional and structural information contained within an entire genome.

Two general types of high-density microarray platforms are used most widely. The first type of microarrays consist of short (up to  $\sim$ 100-mer) oligonucleotide probes, which are synthesized directly on the surface of arrays by photolithography using light-sensitive synthetic chemistry and photolithographic masks, an ink-jet device, or programmable optical mirrors. These types of arrays can be further distinguished based on the type of probes of which they consist. There are the so-called semi-whole-genome (non-tiling) expression arrays that represent only the predicted (annotated) features of a genome, such as exons or splice junctions. On the other hand, the truly whole-genome tiling arrays (hereafter referred to as WGAs) are designed to interrogate an entire genome in an unbiased fashion (1–3). This class of microarrays consists of non-overlapping or partially overlapping probes that are tiled or spaced at regular intervals to cover the entire genome from end to end. The WGAs are already being manufactured with over 6,000,000 discrete features per array, with every feature comprising millions of copies of the specific probe sequence. For instance, the Affymetrix® GeneChip® *Arabidopsis* tiling array is a single array comprised of over 3.2 million perfect match and mismatch probe pairs ( $\sim$ 6.4 million probes total) tiled with 35 base pair spacing throughout the complete non-repetitive portion of the *Arabidopsis thaliana* genome.

The second array platform is made by mechanically printing/spotting probes, such as amplified PCR products, oligonucleotides, or cloned DNA fragments, onto the glass slides (referred to from this point as spotted arrays). Spotted arrays generally have a much lower feature density, usually on the order of approximately 10,000–40,000 spots per chip, than the *in situ* synthesized oligonucleotide arrays. Overall, the high reproducibility, the ability to synthesize probes representing virtually any sequence of a finished genome, and the increased feature density have made the WGAs the preferred platform for whole-genome analysis. Moreover, the ability to utilize relatively short probe lengths combined with the flexibility of using multiple overlapping probes representing every region of an organism's genome makes WGAs an ideal choice for detecting the broadest range of genomic features (including even small polymorphisms and splice variants), rivaled only by the ultradeep sequencing (discussed in this volume in the chapter by Fox et al.). Furthermore, the specificity gained from using short probes also allows repetitive regions or gene family members to be distinguished from one another. Here, we discuss several approaches using WGAs for transcriptome characterization, novel gene discovery, analysis of alternative splicing, mapping of regulatory DNA motifs using chromatin immunoprecipitation (ChIP-chip), methylome analysis, and sequence polymorphism discovery.

---

## 2. Applications of Whole-Genome Tiling Arrays (WGAs)

### 2.1. Using WGAs for Transcriptome Characterization and Gene Discovery

Although computational methods of gene prediction have steadily improved over the past decade, such methods alone still do not enable the accurate determination of the gene structure and/or identify all transcription units in an organism. Additionally, large-scale cloning and sequencing of complementary DNA (cDNA) molecules corresponding to expressed gene products, the traditional approach for identifying coding regions, often misses very low abundance transcripts. Furthermore, any given cDNA collection can be devoid of transcripts that are expressed only in a subset of tissues and/or in response to a specific physiological or environmental condition(s). Hence, the WGAs that cover the entire sequence of the genome of interest represent an attractive alternative that largely circumvents such issues. For example, to study the tissue-specific expression comprehensively, the targets for the WGA hybridization should be generated from a variety of tissues. In brief, total RNA samples from the selected set of tissues are used to make the first strand cDNAs using an oligo(dT) primer containing a linked promoter for T7 RNA polymerase (T7 RNAP), followed by the conversion into the double-stranded form and an *in vitro* transcription by T7 RNAP to generate as well as amplify the biotinylated complementary RNA (cRNA). This protocol, based on a method devised by Eberwine and colleagues (4), results in an unbiased representation of all expressed gene products contained in the total RNA samples, while allowing for an amplification of the targets in sufficient quantity for hybridization to WGAs.

Remarkably, the very first data sets addressing the transcriptional activity in the various tissues in *Arabidopsis* using WGAs identified a large number of novel sites of expression that were missed by computational gene prediction algorithms and cDNA collections (2, 5–7). To define such novel sites of transcriptional activity, the raw data were first pre-processed by dividing the intensity values for each probe by the median intensity value of all probes, including the perfect match (PM), mismatch (MM), and control probes present on the chip, thereby establishing the background noise level in a given experiment. Then, regions of transcriptional activity from the array data that did not correspond to annotated genic units within the most recent genome annotation were classified as novel “expressed” regions if the median intensity value of the probes in that region fell above a certain background cutoff threshold (operationally defined through a metric summarizing the signal emanating from the promoter regions). This approach gives an unbiased tally of novel genic units, which is based solely on the probe intensity values for the regions that fall outside of annotated gene structures.

Interestingly, many of these newly identified transcripts are expressed from the antisense strand relative to previously annotated transcripts (2), and many of them possess an intriguing regulatory potential. For instance, this study has revealed an antisense transcript overlapping the 3' end of the mRNA for the key repressor that regulates flowering time, *FLOWERING LOCUS C* (*FLC*). This antisense transcript may act as a substrate for the biogenesis of the small interfering RNAs (siRNAs) responsible for the heterochromatization and subsequent silencing of this genomic region (8). Additionally, this study has uncovered evidence for the expression in centromeric regions, which were previously thought to be mostly devoid of active transcription (2). Thus, WGAs offer an extremely powerful platform for the discovery of novel transcription units.

## **2.2. Population Genomic Studies Using WGAs**

The genomic content of individuals from the same species can vary in sequence as a result of diverse evolutionary processes. Comprehensive polymorphism data constitute a powerful resource for identifying the sequence variants that affect the phenotypic differences among the individuals (9). Although direct sequencing of individual populations is the most straightforward method for amassing the comprehensive polymorphism data, this methodology has not yet become cost-effective and widely accessible in most organisms (10). To circumvent these problems, Clark et al. (11) applied WGAs for comprehensive polymorphism detection in *Arabidopsis*, expanding upon the strategy used earlier to identify a large fraction of the SNP variation in human and mouse (12, 13). To this end, Clark et al. targeted 19 wild accessions of *A. thaliana*, selected so as to sample the maximal span of genetic diversity. Each DNA sample was whole genome amplified to generate sufficient DNA for hybridization, partially fragmented with DNase I, end-labeled with biotinylated dUTP and ddUTP, and used to probe WGAs spanning the entire *Arabidopsis* genome with single base resolution on both strands, hence interrogating nearly a billion features per experiment.

This WGA-based approach succeeded in capturing much of the common sequence polymorphism found in the worldwide *A. thaliana* population. Furthermore, this data enabled the systematic identification of the types of sequences that differ between accessions, as well as provided a high-resolution map of the genome-wide distribution of polymorphism in this reference plant. Altogether, more than 1 million non-redundant single nucleotide polymorphisms (SNPs) were identified, and ~4% of the genome was identified as being highly dissimilar (or even deleted) relative to the reference (Col-0) genome sequence. Curiously, the patterns of polymorphism between the 19 wild accessions and the reference genome sequence (Col-0) are highly non-random among the gene families, with genes mediating the

interaction with the biotic environment exhibiting an exceptionally high polymorphism levels. Also, regional variation in polymorphism was readily apparent at the chromosome-level scale. This WGA-enabled polymorphism data set provides an unprecedented resource for further evolutionary, genetic, and functional genomic studies.

Two related studies used WGA hybridization of DNA samples from wild accessions of *A. thaliana* to measure the genetic diversity and intraspecific polymorphism between individuals (14, 15). These studies demonstrated that total and pairwise diversity was higher near the centromeres and the heterochromatic knob region, which are highly repetitive in nature and are less active in transcription. Furthermore, the overall diversity between the *Arabidopsis* accessions positively correlated with recombination rate. The combined data from the three studies (11, 14, 15) has enabled the production of an *Arabidopsis* genotyping array, which contains 250,000 SNPs and is commercially available from Affymetrix. This SNP array assures more than adequate coverage for the genome-wide association mapping studies in *Arabidopsis* (15), thus providing the research community with the framework for the future in-depth studies of genetic variation in plants. Taken together, these studies demonstrate that, even in the absence of sequence data for a number of individuals from the same species, population genomic studies can still be carried out successfully using hybridization to WGAs.

### **2.3. ChIP-Chip Studies Using WGAs**

Transcription represents the first major control point in gene expression pathways. Although the overall process of transcription can be regulated by a variety of mechanisms, the most prominent among them are those mediated by the DNA-binding transcription factors and by chromatin structure, which is largely modulated via covalent modifications of the histone N-terminal tails. Chromatin immunoprecipitation (ChIP) with an antibody specific to the protein or modification of interest, followed by the hybridization to WGAs of the DNA extracted from the captured chromatin fragments (i.e., ChIP-chip), has emerged as a powerful approach for gaining insight into the genome-wide distribution of the specific transcription factors or histone modifications ((16–20) and the chapter by Morohashi et al. in this volume). The quality of the antibody used in immunoprecipitation of the DNA-bound protein of interest is the major limiting factor of this technique, because it is critical for achieving an effective enrichment of the protein-bound DNA fragments for hybridization to WGAs.

In one instructive study of this kind, the antibodies against the sequence-specific transcription factor TGA2 were used to map its binding sites genome-wide after the treatment of *Arabidopsis* plants with the phytohormone salicylic acid (SA) (21).

The TGA2-crosslinked, immunoprecipitated DNA fragments were nonspecifically amplified to obtain enough material for the hybridization, fragmented with DNase I, end-labeled with biotinylated-ddATP using terminal transferase, and hybridized to two distinct types of WGAs. The first platform contained 190,000 probes representing 2 kb regions upstream of all annotated genes at a density of seven probes per promoter, while the second platform represented the entire *Arabidopsis* genome at a density of one probe per 90 bases. This study revealed 51 putative binding sites for TGA2, including the only previously identified (in the promoter of *At2g14610 PR-1* gene), as well as 15 putative binding sites that lie outside of presumed promoter regions. Additionally, when the effect of SA treatment on gene expression was measured using standard gene expression arrays, SA-induced transcripts were found to be significantly overrepresented among the genes neighboring the putative TGA2-binding sites. This example illustrates how the combined use of WGA platforms for ChIP-chip and gene expression studies can give important clues as to how sequence-specific transcription factors govern the key regulatory networks within plant cells.

Covalent modification of histones is another key mechanism controlling the eukaryotic genome dynamics. Motivated in part by the evidence that tri-methylation of lysine 27 of histone H3 (H3K27me3) plays critical roles in regulating development in animals (16, 22, 23) and plants (24–27), WGA-based profiling of the H3K27me3 in *Arabidopsis* was undertaken (28, 29). These analyses revealed, for the first time, that H3K27me3 is a major silencing mechanism in *Arabidopsis* that regulates an unexpectedly large number of genes located in mostly euchromatic regions. Furthermore, analysis of the H3K27me3 profiles in the relevant mutant backgrounds suggested that establishment and maintenance of this histone modification is largely independent of other epigenetic pathways, such as DNA methylation or RNA silencing. Interestingly, the genomic domains marked by H3K27me3 associate almost exclusively and co-extensively with binding sites for TERMINAL FLOWER 2/LIKE HETEROCHROMATIN PROTEIN 1, which is similar to the HETEROCHROMATIN PROTEIN 1 (HP1) of metazoans and *Schizosaccharomyces pombe* (28, 29). However, the genome-wide distribution of H3K27me3 was unaffected in *lhp1* mutant, suggesting that TFL2/LHP1 is not involved in the deposition of this chromatin modification but rather is a part of the epigenetic mechanism that represses the expression of genes that are marked with the H3K27me3. Therefore, ChIP-chip experiments with WGAs can be very powerful in revealing the key regulatory mechanisms controlling the complex dynamics of the genome activity in plants. As the number of WGA-based ChIP-chip experiments grows, a much more complete view of the transcriptional networks controlling plant growth and development will emerge.

## 2.4. Characterization of the Methylome Using WGAs

DNA methylation is a conserved epigenetic silencing mechanism involved in many important biological phenomena, including defense against transposon proliferation, genomic imprinting, and regulation of gene expression. DNA methylation is a heritable epigenetic modification that is perpetuated through DNA replication by DNA methyltransferases (30, 31). DNA methylation allows to regulate the expression of a number of coding regions without mutation to the DNA sequence, and it can occur in *cis* (i.e., the gene itself is methylated) or in *trans* (when the methylation event at another site in the genome regulates the target gene) (32–34).

In *Arabidopsis*, DNA methylation is established in all sequence contexts by DRM1/2, which are homologs of the mammalian DNMT3a/b de novo DNA methyltransferases (35, 36). DRM1/2 activity can be directed to a precise genomic location by RNA-directed DNA methylation (RdDM) system that involves 21–24 nucleotide small RNA (smRNA) generated in a DICER-LIKE3-dependent manner and acting in concert with ARGONAUTE4 (37–39). On the other hand, DNA methylation within the context of CpG dinucleotide is stably maintained maintained through genome replication by the DNA methyltransferase MET1, a homolog of mammalian DNA methyltransferase1 (40–42). Finally, the plant-specific DNA methyltransferase CMT3 primarily targets cytosines in the CHG sequence context (where H = A, C, T) (43).

WGAs allow to comprehensively map the methylome, i.e., the sum total of the sites of DNA methylation within the *Arabidopsis* genome (3, 44–46). In the pioneering study of this kind, an antibody against the 5-methyl cytosine was used to generate the target for interrogating the WGAs spanning the entire *Arabidopsis* genome (3). The resulting DNA methylation map reveals that approximately 19% of the genome is methylated, with the regions containing the highest density of methylation located in highly repetitive regions of the genome, such as centromeric heterochromatin. Predictably, the highest levels of methylation were seen in pseudogenes and unexpressed genes, but surprisingly, a considerable amount of methylation was distributed in euchromatin. However, only ~5% of expressed genes contained methylation upstream of their ORFs (promoters), while 33% of the transcribed regions of these genes were methylated (body methylation), consistent with an earlier smaller-scale study (47). Another surprising discovery from these WGA studies was that most of the genes that contain DNA methylation within their transcribed regions are highly expressed and constitutively active. Furthermore, the distribution of DNA methylation is clearly different between transposons and genes: while DNA methylation of transposons is distributed across their entire length, methylation density in genes was low in the promoter regions, but gradually increased within the transcribed region and dropped off again in the 3'

flanking sequences. This pattern may indicate negative selection against methylating the 5' and 3' ends of expressed genes, e.g., because of incompatibility with transcription initiation and termination.

The methyl groups in DNA are not static but can be removed by the DNA demethylases (48–51). *Arabidopsis* has four such DNA demethylases, REPRESSOR OF SILENCING1 (ROS1), DEMETER (DME), DEMETER-LIKE2 (DML2), and DEMETER-LIKE3 (DML3) (48, 49, 51). DME is required for genomic imprinting during *Arabidopsis* embryo development (52), while the closely related ROS1 is involved in transcriptional silencing of a transgene (51). WGAs were employed to globally map the sites of DNA demethylation within the *Arabidopsis* genome, via comparing the methylome in WT and mutant plants lacking three of the DNA demethylases (ROS1, DML2, DML3) (53). It appears that 179 loci are actively demethylated by one or all of these enzymes in *Arabidopsis*, and interestingly, demethylation in the coding regions primarily occurs at both the 5' and 3' ends, i.e., in a pattern opposite to the overall distribution of DNA methylation. This suggests that DNA methylation is highly dynamic and that the process of demethylation may act to protect the genes from potentially deleterious methylation events. Taken together, these first methylome studies provide important insights into the nature as well as the function of this important epigenetic mark.

---

### 3. Technical Considerations Regarding the WGA Analyses

#### 3.1. Sequence-Specific Probe Effects

Although tiling microarrays are very powerful, as illustrated in the preceding sections, several limitations and important technical considerations must be taken into account. One major technical constraint that is inherent to the concept of WGAs lies in the severely limited freedom of choice in designing the probes. This limitation translates into the widespread sequence-based probe effects. For example, ~20% of probes located entirely within a known (i.e., experimentally proven) exon exhibit twofold or higher difference in the signal intensity relative to the average intensity of their two neighboring probes located within same exon (54). While in principle such probe behavior can result from alternative splicing or from cross-hybridization to other transcribed sequences that map to unrelated genomic locations, in reality the most significant source of such effects is the variability in thermodynamic properties of the probes themselves, as dictated by their respective sequences.

Several approaches have been used in an attempt to correct for such unevenness in the probe behavior. One alternative is to tackle the problem early, i.e., at the stage of the design of the array. For instance, maskless tiling arrays manufactured by NimbleGen are composed entirely of isothermal probes, whereby the length of each molecule is varied in order to attain a consistent melting temperature (usually set at 76°C). Although the isothermal arrays should produce uniform probe behavior, this advantage comes at a price of the decrease in the feature density as compared to the one afforded by the Affymetrix platform, as well as reduced resolution. Furthermore, while the theoretical design of isothermal probes is typically based on the nearest-neighbor behavior of the respective oligonucleotides in solution, in practice the behavior of the array probes is strongly influenced by additional factors, such as steric hindrance on the microarray surface, probe–probe interaction, and secondary structure formation (55, 56).

An alternative strategy to address the unevenness in the probe behavior relies on statistical methods. Such approaches extend the earlier efforts to model the sequence-specific probe behavior for gene expression microarrays (57, 58). For example, MAT (model-based analysis of tiling arrays) predicts the baseline probe behavior by considering the 25-mer sequence and copy number of all probes on a single Affymetrix tiling array (59). This approach standardizes the probe value through the probe model, eliminating the need for sample normalization. As opposed to estimating probe behavior from multiple samples (60–62), MAT can standardize the signals of each probe in each array individually. MAT approaches perform particularly well in ChIP-chip applications that measure the genome-wide transcription factor binding and can detect with high accuracy the enriched regions from a single or multiple ChIP samples. This is due to the fact that most probes in ChIP-chip analyses measure only unspecific binding, because transcription factors usually bind only to a small fraction of the genome. One variation on the MAT theme is to use an *a priori* sequence-dependent physical model of probe-specific intensity bias (occurring primarily due to unspecific binding), instead of estimating it from the data (63).

Finally, the third alternative, which may hold a particular appeal to experimentalists, is to empirically calibrate the behavior of the probes on the array against a suitable reference sample. For example, in the global mapping study of transcriptional activity in yeast, non-equivalencies in the probe behavior were corrected via experimental RNA/DNA hybridization-based model, by correcting for the background as well as adjusting the signal of each probe by sequence-specific parameters, which was estimated from a calibration set of genomic DNA hybridizations (64). The following normalization methods were evaluated: (1) dividing RNA signal by DNA signal

and then taking base 2 logarithm; (2) background-subtracting the RNA signal, dividing it by DNA signal, then applying variance stabilizing normalization (vsn, log base 2); and (3) in addition to method 2, dropping the 5% weakest probes in the DNA hybridization. Method 3 yielded the highest gain in signal to noise ratio, which was estimated as follows. Noise was estimated from the median of absolute differences between each pair of probes on the Crick strand of chromosome IV whose start points were three intermediate probes apart. Signal was obtained from the difference between 99% and 1% quantiles of all these probes. The optimal method of normalization for the individual probe behavior increased the signal/noise ratio on average by 1.7-fold.

### ***3.2. Distinguishing the Signal from Noise***

Although the issue of distinguishing the signal from noise is not unique to the whole-genome microarrays in particular, it is particularly important in the case of WGAs, because as opposed to all other types of microarrays analyses, WGAs assume no underlying gene models or annotations. Hence, a radically different strategy is required to make the decisions on how to make the “present” calls (54). In the early studies, positive probes were called based on a probe signal cutoff (65), and the genomic regions containing a significant number of positive probes were designated as transfrags (transcribed fragments). At present, two major alternative strategies to identify the regions of significant signal on WGAs are based on either the sliding window approaches or on structured change point detection algorithm (66, 67). The latter approach aims to segment the genome using dynamic programming in an unbiased fashion into regions with different expression levels in such a way that the probe signals are similar within each region. Such methods are reported to give more accurate estimates of change point locations, as well as depend on fewer user-defined parameters (64). However, the sliding window-based approaches remain more common.

The authors of this chapter have experience with TileMap (60). This package was originally developed for ChIP-chip analysis, but it can be used to analyze other types of genome-wide data, such as that of the entire transcriptome or methylome. The distinctive feature of TileMap is that it treats every probe as a separate entity, rather than computing a metric for a particular gene. Therefore, rather than generating a gene-level measurement of intensity changes, TileMap enables an unbiased identification of those genomic regions that demonstrate significantly up- or down-regulated hybridization between two different sets of arrays. An additional advantage of this algorithm is that it goes beyond the ability of just making the “present” calls, but rather allows complex multiple-condition comparisons (e.g., mutant 1 > WT > mutant 2).

In the first step of the TileMap procedure, a t-like test statistic is computed separately for each probe on the array, using a hierarchical empirical Bayes model to pool information from all probes across the array. On the other hand, during the calculation of the conventional t-statistic, only the estimate of the probe's own standard deviation is accounted for. This significantly increases the sensitivity of the analysis in the very common circumstances when there are only a small number (2–3) of replicates available for each condition. In the second step, the test statistics of probes within a genomic region are used to infer whether the region is of interest or not (i.e., whether it shows transcriptional activities of interest). TileMap offers two different ways to do this: users can choose to combine neighboring probes by using either a hidden Markov model (HMM) or a moving average method (MA). Finally, TileMap uses unbalanced mixture subtraction (UMS) to provide approximate local false discovery rate (lfdr) estimates for MA and model parameters for HMM. Compared with the commonly employed permutation test, UMS performs better for complex multiple-sample comparisons, such as mutant 1 > WT > mutant 2. Importantly, while UMS estimates the lfdr for a null hypothesis  $H_0$ : “not (mutant 1 > WT > mutant 2)”, permutation test usually can only provide lfdr for a null  $H_0$ : “mutant 1 = WT = mutant 2”.

### ***3.3. Primary vs. Secondary Effects in Tiling Microarray Experiments***

One cautionary point is warranted concerning the widespread practice of using microarrays to reveal the expression changes in various mutant backgrounds compared to WT. In the case of constitutive mutants, the differentially expressed regions represent the sum of primary and secondary effects of inactivating the respective cellular factor. Distinguishing the direct from secondary consequences of a given mutation on the transcriptome can be challenging and requires special consideration during the design stage and/or extensive follow-up experimental and/or bioinformatic analyses. While this problem is by no means unique to WGA-based studies, it can become particularly acute in this case because of the sheer volume of data that such studies generate.

While there is no single universal solution to this problem, several considerations may be helpful. One of these concerns the analyses of transcription factors, which often tend to regulate other transcription factors, forming branched networks. For the sake of example, if several sets of genes (regulons) are coordinately affected upon inactivating the factor X, one can query the public microarray repositories (e.g., [www.weigelworld.org/resources/microarray/AtGenExpress](http://www.weigelworld.org/resources/microarray/AtGenExpress)) and/or specialized transcription factor databases (e.g., AGRIS, [arabidopsis.med.ohio-state.edu](http://arabidopsis.med.ohio-state.edu)) for the transcription factor(s) that may directly control the expression of these gene sets. A reasonable hypothesis then would be that the effect of X on these otherwise disparate gene sets is indirect and mediated by its regulating the expression of these transcription factor(s).

A more radical approach to the problem of secondary effects would be to attempt to bypass this issue altogether. In one example, this was achieved by putting the transcription factor under study under exclusively posttranslational control, via fusing it to the rat glucocorticoid receptor (GR). Simultaneous treatment of the transgenic line expressing such a chimeric factor by the activating glucocorticoid dexamethasone and the translational inhibitor cycloheximide then led to the transcriptional induction of the direct target genes only, while the expression of any secondary effects was blocked (68). However, the most generally applicable tools allowing to filter out secondary effects are conditional mutants. In this case, one can apply the restrictive condition at will and monitor the real-time progression of the ensuing changes in the transcriptome by microarray analysis of the early timepoints of the response. In this case, the expectation is that the inactivation of the transcription factor should have a very rapid effect on its immediate target promoters, comparable with inactivation of general transcription. On the other hand, a considerably longer period of time would be required to develop secondary effects, because such effects must be preceded by significant alterations in the mRNA as well as in the protein levels of the immediate downstream targets of the transcription factor in question.

Unfortunately, temperature-sensitive mutations, which are widely used for this type of analysis in microorganisms, are rather rare in plants. However, the authors have been successful in implementing an inducible RNAi (iRNAi) for creating a conditional knockdown of the subunits of the exosome complex in *Arabidopsis*. The exosome is an essential and conserved RNA-degrading and RNA-processing complex that has multiple and diverse RNA targets that are yet to be comprehensively defined in any eukaryote. An iRNAi system was engineered by expressing the constructs containing the segments of the exosome complex subunits *RRP4* or *RRP41* as a pair of inverted repeats separated by an intron, under the control of an estradiol-regulated chimeric trans-activator XVE (69). Growing such exosome iRNAi plants on estradiol-containing media induced the RNAi-mediated knockdown of *RRP4* (*rrp4*<sup>iRNAi</sup>) or *RRP41* (*rrp41*<sup>iRNAi</sup>) mRNA, resulting in the growth arrest and subsequent death of seedlings. Importantly, growth arrest was preceded by the highly specific molecular phenotype associated with the defective processing of the 5.8S ribosomal RNA, which is highly specific to exosome malfunction (70), and is never observed in WT plants exposed to estradiol (neither is growth inhibition). Thus validated conditional iRNAi knockdown system was subsequently used in conjunction with WGAs to comprehensively define the *Arabidopsis* exosome targets (71). In contrast, WGA analysis of a constitutive loss of function mutant of one of the subunits of this complex

produced massive amounts of secondary effects, even though this particular mutant had little if any phenotype at the whole-plant level (71).

### **3.4. Inhibition of RNA Degrading/Processing Enzymes as a General Strategy to Uncover the Hidden Dynamics in the Transcriptome**

Numerous studies during the past few years revealed the existence of the vast “dark matter” in eukaryotic transcriptomes, in the form of noncoding (nc) RNAs with unknown function (72). Although there is much debate as to what fraction of these ncRNAs have biological targets vs. merely represent spurious transcriptional activity (73), it is important to comprehensively understand such events. In this regard, it is instructive to consider the finding of the “deeply hidden” layer in the *Arabidopsis* transcriptome that is only detectable under the conditions of exosome knockdown (71). Such RNAs are largely composed of intergenic noncoding transcripts that emanate from the repetitive, heterochromatic regions of the genome. Apparently, these transcripts are tightly down-regulated by the constitutive exosome activity to the extent that they are virtually undetectable under normal conditions. On a more general note, the exosome is but one among many diverse RNA processing/degrading activities in the cell, and hence a logical extension of this strategy would be to undertake a systematic identification and categorization of the transcriptome-wide consequences of modulating the activities of a wide variety RNA decay and processing factors.

While in hindsight this approach may seem intuitive, in practice such analyses have been conducted only rarely, and mostly in the context of specialized studies focusing on specific class of transcripts and/or specific aspects of RNA metabolism. In one such example, He et al. used microarrays to investigate the RNAs in *Saccharomyces cerevisiae* that are stabilized upon mutating the key components of the nonsense-mediated mRNA decay (NMD) pathway – a specialized mechanism dedicated to the degradation of mRNAs containing the premature stop codons (74). In another approach to identify the transcripts directly regulated by NMD, the same group examined which RNAs become rapidly down-regulated upon restituting the NMD pathway in the NMD-defective cells, using conditional promoter (75). These combined studies succeeded in defining a near-comprehensive core set of cellular transcripts regulated by NMD, many of which have not been previously known, e.g., such as those RNAs that fail to splice and escape into the cytoplasm, mRNAs with abnormally long 3' UTRs, mRNAs with upstream open reading frames, mRNAs that are subject to leaky scanning resulting in the use of out-of-frame initiator codons, mRNAs translated via +1 frameshifting, bicistronic mRNAs, transcripts encoded by pseudogenes, as well as those emanating from the transposable elements or from their LTR sequences. In another study, a conditional promoter was used to inhibit the expression of an essential subunit of the nuclear RNase

P in yeast, combined with the WGA-based monitoring of the ensuing changes in the transcriptome during the time course of depletion (76). This study led to the discovery of 73 novel ncRNAs, many of them antisense relative to the previously annotated ORFs – a surprisingly large number for the best-annotated eukaryotic genome. We therefore propose that manipulating the activities of the key factors of plant RNA metabolism may be a productive approach for mining the depths of the dynamic plant transcriptomes.

---

## Acknowledgments

Research on plant gene expression in the Belostotsky lab is supported by USDA (award 2007-35301-18207), the National Science Foundation (awards MCB-0424651 and DBI-0724168), and BARD (award 3756-05). B.D.G. is a Damon Runyon Fellow, supported by the Damon Runyon Cancer Research Foundation.

## References

1. Penterman, J., Zilberman, D., Huh, J.H., Ballinger, T., Henikoff, S., and Fischer, R.L. (2007) DNA demethylation in the *Arabidopsis* genome. *Proc. Natl. Acad. Sci. USA* **104**, 6752–6757.
2. Kakutani, T. (2002) Epi-alleles in plants: inheritance of epigenetic information over generations. *Plant Cell Physiol.* **43**, 1106–1111.
3. Kankel, M.W., Ramsey, D.E., Stokes, T.L., Flowers, S.K., Haag, J.R., Jeddeloh, J.A., Riddle, N.C., Verbsky, M.L., and Richards, E.J. (2003) *Arabidopsis* MET1 cytosine methyltransferase mutants. *Genetics* **163**, 1109–1122.
4. Clark, R.M., Schweikert, G., Toomajian, C., Ossowski, S., Zeller, G., Shinn, P., Warthmann, N., Hu, T.T., Fu, G., Hinds, D.A., Chen, H., Frazer, K.A., Huson, D.H., Scholkopf, B., Nordborg, M., Ratsch, G., Ecker, J.R., and Weigel, D. (2007) Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science* **317**, 338–342.
5. Van Gelder, R.N., von Zastrow, M.E., Yool, A., Dement, W.C., Barchas, J.D., and Eberwine, J.H. (1990) Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc. Natl. Acad. Sci. USA* **87**, 1663–1667.
6. Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S.W., Chen, H., Henderson, I.R., Shinn, P., Pellegrini, M., Jacobsen, S.E., and Ecker, J.R. (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in *arabidopsis*. *Cell* **126**, 1189–1201.
7. Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J.J., Goldberg, R.B., Jacobsen, S.E., and Fischer, R.L. (2002) DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *arabidopsis*. *Cell* **110**, 33–42.
8. Hanada, K., Zhang, X., Borevitz, J.O., Li, W.H., and Shiu, S.H. (2007) A large number of novel coding small open reading frames in the intergenic regions of the *Arabidopsis thaliana* genome are transcribed and/or under purifying selection. *Genome Res.* **17**, 632–640.
9. Daruwala, R.S., Rudra, A., Ostrer, H., Lucito, R., Wigler, M., and Mishra, B. (2004) A versatile statistical analysis algorithm to detect genome copy number variation. *Proc. Natl. Acad. Sci. USA* **101**, 16292–16297.
10. Keles, S., van der Laan, M.J., Dudoit, S., and Cawley, S.E. (2006) Multiple testing methods for ChIP-Chip high density

oligonucleotide array data. *J. Comput. Biol.* **13**, 579–613.

11. Kim, H., Snesrud, E.C., Haas, B., Cheung, F., Town, C.D., and Quackenbush, J. (2003) Gene expression analyses of *Arabidopsis* chromosome 2 using a genomic DNA amplicon microarray. *Genome Res.* **13**, 327–340.
12. Hekstra, D., Taussig, A.R., Magnasco, M., and Naeff, F. (2003) Absolute mRNA concentrations from sequence-specific calibration of oligonucleotide arrays. *Nucleic Acids Res.* **31**, 1962–1968.
13. Zhang, X., Clarenz, O., Cokus, S., Bernatavichute, Y.V., Pellegrini, M., Goodrich, J., and Jacobsen, S.E. (2007) Whole-genome analysis of histone H3 lysine 27 trimethylation in *Arabidopsis*. *PLoS Biol.* **5**, e129.
14. Sablowski, R.W. and Meyerowitz, E.M. (1998) A homolog of NO APICAL MERISTEM is an immediate target of the floral homeotic genes APETALA3/PISTILLATA. *Cell* **92**, 93–103.
15. Tran, R.K., Henikoff, J.G., Zilberman, D., Ditt, R.F., Jacobsen, S.E., and Henikoff, S. (2005) DNA methylation profiling identifies CG methylation clusters in *Arabidopsis* genes. *Curr. Biol.* **15**, 154–159.
16. Cao, X. and Jacobsen, S.E. (2002) Role of the *Arabidopsis* DRM methyltransferases in de novo DNA methylation and gene silencing. *Curr. Biol.* **12**, 1138–1144.
17. Wu, J., Smith, L.T., Plass, C., and Huang, T.H. (2006) ChIP-chip comes of age for genome-wide functional analysis. *Cancer Res.* **66**, 6899–6902.
18. Turck, F., Roudier, F., Farrona, S., Martin-Magniette, M.L., Guillaume, E., Buisine, N., Gagnot, S., Martienssen, R.A., Coupland, G., and Colot, V. (2007) *Arabidopsis* TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. *PLoS Genet.* **3**, e86.
19. Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K., Koseki, H., Fuchikami, T., Abe, K., Murray, H.L., Zucker, J.P., Yuan, B., Bell, G.W., Herbolsheimer, E., Hannett, N.M., Sun, K., Odom, D.T., Otte, A.P., Volkert, T.L., Bartel, D.P., Melton, D.A., Gifford, D.K., Jaenisch, R., and Young, R.A. (2006) Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**, 301–313.
20. Picard, F., Robin, S., Lavielle, M., Vaisse, C., and Daudin, J.J. (2005) A statistical approach for array CGH data analysis. *BMC Bioinformatics* **6**, 27.
21. Johnson, J.M., Edwards, S., Shoemaker, D., and Schadt, E.E. (2005) Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments. *Trends Genet.* **21**, 93–102.
22. Bertone, P., Stolc, V., Royce, T.E., Rozowsky, J.S., Urban, A.E., Zhu, X., Rinn, J.L., Tongprasit, W., Samanta, M., Weissman, S., Gerstein, M., and Snyder, M. (2004) Global identification of human transcribed sequences with genome tiling arrays. *Science* **306**, 2242–2246.
23. Zilberman, D., Cao, X., Johansen, L.K., Xie, Z., Carrington, J.E., and Jacobsen, S.E. (2004) Role of *Arabidopsis* ARGONAUTE4 in RNA-directed DNA methylation triggered by inverted repeats. *Curr. Biol.* **14**, 1214–1220.
24. Shiu, S.H. and Borevitz, J.O. (2008) The next generation of microarray research: applications in evolutionary and ecological genomics. *Heredity* **100**, 141–149.
25. Chanvivattana, Y., Bishopp, A., Schubert, D., Stock, C., Moon, Y.H., Sung, Z.R., and Goodrich, J. (2004) Interaction of Polycomb-group proteins controlling flowering in *Arabidopsis*. *Development* **131**, 5263–5276.
26. David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C.J., Bofkin, L., Jones, T., Davis, R.W., and Steinmetz, L.M. (2006) A high-resolution map of transcription in the yeast genome. *Proc. Natl. Acad. Sci. USA* **103**, 5320–5325.
27. Li, W., Meyer, C.A., and Liu, X.S. (2005) A hidden Markov model for analyzing ChIP-chip experiments on genome tiling arrays and its application to p53 binding sequences. *Bioinformatics* **21** Suppl 1, i274–i282.
28. Stolc, V., Samanta, M.P., Tongprasit, W., Sethi, H., Liang, S., Nelson, D.C., Hegeman, A., Nelson, C., Rancour, D., Bednarek, S., Ulrich, E.L., Zhao, Q., Wrobel, R.L., Newman, C.S., Fox, B.G., Phillips, Jr., G.N., Markley, J.L., and Sussman, M.R. (2005) Identification of transcribed sequences in *Arabidopsis* thaliana by using high-resolution genome tiling arrays. *Proc. Natl. Acad. Sci. USA* **102**, 4453–4458.
29. Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T., and Henikoff, S. (2007)

Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nat. Genet.* **39**, 61–69.

30. Zhu, J., Kapoor, A., Sridhar, V.V., Agius, F., and Zhu, J.K. (2007) The DNA glycosylase/lyase ROS1 functions in pruning DNA methylation patterns in *Arabidopsis*. *Curr. Biol.* **17**, 54–59.
31. Stam, M. and Mittelsten Scheid, O. (2005) Paramutation: an encounter leaving a lasting impression. *Trends Plant Sci.* **10**, 283–290.
32. Johnson, W.E., Li, W., Meyer, C.A., Gottardo, R., Carroll, J.S., Brown, M., and Liu, X.S. (2006) Model-based analysis of tiling-arrays for ChIP-chip. *Proc. Natl. Acad. Sci. USA* **103**, 12457–12462.
33. Li, C.F., Pontes, O., El-Shami, M., Henderson, I.R., Bernatavichute, Y.V., Chan, S.W., Lagrange, T., Pikaard, C.S., and Jacobsen, S.E. (2006) An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in *Arabidopsis thaliana*. *Cell* **126**, 93–106.
34. Agius, F., Kapoor, A., and Zhu, J.K. (2006) Role of the *Arabidopsis* DNA glycosylase/lyase ROS1 in active DNA demethylation. *Proc. Natl. Acad. Sci. USA* **103**, 11796–11801.
35. Swiezewski, S., Creville, P., Liu, F., Ecker, J.R., Jerzmanowski, A., and Dean, C. (2007) Small RNA-mediated chromatin silencing directed to the 3' region of the *Arabidopsis* gene encoding the developmental regulator, FLC. *Proc. Natl. Acad. Sci. USA* **104**, 3633–3638.
36. Bickel, K.S. and Morris, D.R. (2006) Silencing the transcriptome's dark matter: mechanisms for suppressing translation of intergenic transcripts. *Mol. Cell* **22**, 309–316.
37. Chung, H.R., Kostka, D., and Vingron, M. (2007) A physical model for tiling array analysis. *Bioinformatics* **23**, i80–i86.
38. Hinds, D.A., Stuve, L.L., Nilsen, G.B., Halperin, E., Eskin, E., Ballinger, D.G., Frazer, K.A., and Cox, D.R. (2005) Whole-genome patterns of common DNA variation in three human populations. *Science* **307**, 1072–1079.
39. Cao, X., Aufsatz, W., Zilberman, D., Mette, M.F., Huang, M.S., Matzke, M., and Jacobsen, S.E. (2003) Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. *Curr. Biol.* **13**, 2212–2217.
40. He, F., Li, X., Spatrick, P., Casillo, R., Dong, S., and Jacobson, A. (2003) Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5' to 3' mRNA decay pathways in yeast. *Mol. Cell.* **12**, 1439–1452.
41. Kim, S., Plagnol, V., Hu, T.T., Toomajian, C., Clark, R.M., Ossowski, S., Ecker, J.R., Weigel, D., and Nordborg, M. (2007) Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nat. Genet.* **39**, 1151–1155.
42. Finnegan, E.J. and Dennis, E.S. (1993) Isolation and identification by sequence homology of a putative cytosine methyltransferase from *Arabidopsis thaliana*. *Nucleic Acids Res.* **21**, 2383–2388.
43. Thibaud-Nissen, F., Wu, H., Richmond, T., Redman, J.C., Johnson, C., Green, R., Arias, J., and Town, C.D. (2006) Development of *Arabidopsis* whole-genome microarrays and their application to the discovery of binding sites for the TGA2 transcription factor in salicylic acid-treated plants. *Plant J.* **47**, 152–162.
44. Lippman, Z., Gendrel, A.V., Colot, V., and Martienssen, R. (2005) Profiling DNA methylation patterns using genomic tiling microarrays. *Nat. Methods* **2**, 219–224.
45. Martienssen, R.A., Doerge, R.W., and Colot, V. (2005) Epigenomic mapping in *Arabidopsis* using tiling microarrays. *Chromosome Res.* **13**, 299–308.
46. Yamada, K., Lim, J., Dale, J.M., Chen, H., Shinn, P., Palm, C.J., Southwick, A.M., Wu, H.C., Kim, C., Nguyen, M., Pham, P., Cheuk, R., Karlin-Newmann, G., Liu, S.X., Lam, B., Sakano, H., Wu, T., Yu, G., Miranda, M., Quach, H.L., Tripp, M., Chang, C.H., Lee, J.M., Toriumi, M., Chan, M.M., Tang, C.C., Onodera, C.S., Deng, J.M., Akiyama, K., Ansari, Y., Arakawa, T., Banh, J., Banno, F., Bowser, L., Brooks, S., Carninci, P., Chao, Q., Choy, N., Enju, A., Goldsmith, A.D., Gurjal, M., Hansen, N.F., Hayashizaki, Y., Johnson-Hopson, C., Hsuan, V.W., Iida, K., Karnes, M., Khan, S., Koesema, E., Ishida, J., Jiang, P.X., Jones, T., Kawai, J., Kamiya, A., Meyers, C., Nakajima, M., Narusaka, M., Seki, M., Sakurai, T., Satou, M., Tamse, R., Vaysberg, M., Wallender, E.K., Wong, C., Yamamura, Y., Yuan, S., Shinozaki, K., Davis, R.W., Theologis, A., and Ecker, J.R. (2003) Empirical analysis of transcriptional activity in the *Arabidopsis* genome. *Science* **302**, 842–846.

47. Chekanova, J.A., Gregory, B.D., Reverdatto, S.V., Chen, H., Kumar, R., Hooker, T., Yazaki, J., Li, P., Skiba, N., Peng, Q., Alonso, J., Brukhin, V., Grossniklaus, U., Ecker, J.R., and Belostotsky, D.A. (2007) Genome-wide high-resolution mapping of exosome substrates reveals hidden features in the *Arabidopsis* transcriptome. *Cell* **131**, 1340–1353.

48. Samanta, M.P., Tongprasit, W., Sethi, H., Chin, C.S., and Stolc, V. (2006) Global identification of noncoding RNAs in *Saccharomyces cerevisiae* by modulating an essential RNA processing pathway. *Proc. Natl. Acad. Sci. USA* **103**, 4192–4197.

49. Wu, Z. and Irizarry, R.A. (2005) Stochastic models inspired by hybridization theory for short oligonucleotide arrays. *J. Comput. Biol.* **12**, 882–893.

50. Kinoshita, T., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (2001) Polycomb repression of flowering during early plant development. *Proc. Natl. Acad. Sci. USA* **98**, 14156–14161.

51. Saze, H., Mittelsten Scheid, O., and Paszkowski, J. (2003) Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. *Nat. Genet.* **34**, 65–69.

52. Xiao, W., Gehring, M., Choi, Y., Margossian, L., Pu, H., Harada, J.J., Goldberg, R.B., Pennell, R.I., and Fischer, R.L. (2003) Imprinting of the MEA Polycomb gene is controlled by antagonism between MET1 methyltransferase and DME glycosylase. *Dev. Cell* **5**, 891–901.

53. Zuo, J., Niu, Q.W., and Chua, N.H. (2000) Technical advance: an estrogen receptor-based transactivator XVE mediates highly inducible gene expression in transgenic plants. *Plant J.* **24**, 265–273.

54. Li, L., Wang, X., Sasidharan, R., Stolc, V., Deng, W., He, H., Korbel, J., Chen, X., Tongprasit, W., Ronald, P., Chen, R., Gerstein, M., and Wang Deng, X. (2007) Global identification and characterization of transcriptionally active regions in the rice genome. *PLoS ONE* **2**, e294.

55. Hudson, M.E. and Snyder, M. (2006) High-throughput methods of regulatory element discovery. *Biotechniques* **41**, 673, 675, 677 passim.

56. Schubert, D., Clarenz, O., and Goodrich, J. (2005) Epigenetic control of plant development by Polycomb-group proteins. *Curr. Opin. Plant Biol.* **8**, 553–561.

57. Bulyk, M.L. (2006) DNA microarray technologies for measuring protein–DNA interactions. *Curr. Opin. Biotechnol.* **17**, 422–430.

58. Shchepinov, M.S., Case-Green, S.C., and Southern, E.M. (1997) Steric factors influencing hybridisation of nucleic acids to oligonucleotide arrays. *Nucleic Acids Res.* **25**, 1155–1161.

59. Kapranov, P., Cawley, S.E., Drenkow, J., Bekiranov, S., Strausberg, R.L., Fodor, S.P., and Gingeras, T.R. (2002) Large-scale transcriptional activity in chromosomes 21 and 22. *Science* **296**, 916–919.

60. Borevitz, J.O., Hazen, S.P., Michael, T.P., Morris, G.P., Baxter, I.R., Hu, T.T., Chen, H., Werner, J.D., Nordborg, M., Salt, D.E., Kay, S.A., Chory, J., Weigel, D., Jones, J.D., and Ecker, J.R. (2007) Genome-wide patterns of single-feature polymorphism in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **104**, 12057–12062.

61. Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., Bell, G.W., Otte, A.P., Vidal, M., Gifford, D.K., Young, R.A., and Jaenisch, R. (2006) Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349–353.

62. Grewal, S.I. and Klar, A.J. (1996) Chromosomal inheritance of epigenetic states in fission yeast during mitosis and meiosis. *Cell* **86**, 95–101.

63. Patil, N., Berno, A.J., Hinds, D.A., Barrett, W.A., Doshi, J.M., Hacker, C.R., Kautzer, C.R., Lee, D.H., Marjoribanks, C., McDonough, D.P., Nguyen, B.T., Norris, M.C., Sheehan, J.B., Shen, N., Stern, D., Stokowski, R.P., Thomas, D.J., Trulson, M.O., Vyas, K.R., Frazer, K.A., Fodor, S.P., and Cox, D.R. (2001) Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* **294**, 1719–1723.

64. Ji, H. and Wong, W.H. (2005) TileMap: create chromosomal map of tiling array hybridizations. *Bioinformatics* **21**, 3629–3636.

65. Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S.L., and Lander, E.S. (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326.

66. Gehring, M., Huh, J.H., Hsieh, T.F., Penterman, J., Choi, Y., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (2006) DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting

by allele-specific demethylation. *Cell* **124**, 495–506.

67. Southern, E., Mir, K., and Shchepinov, M. (1999) Molecular interactions on microarrays. *Nat. Genet.* **21**, 5–9.

68. Qi, Y., He, X., Wang, X.J., Kohany, O., Jurka, J., and Hannon, G.J. (2006) Distinct catalytic and non-catalytic roles of ARGO-NAUTE4 in RNA-directed DNA methylation. *Nature* **443**, 1008–1012.

69. Jackson, J.P., Lindroth, A.M., Cao, X., and Jacobsen, S.E. (2002) Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* **416**, 556–560.

70. Alleman, M. and Doctor, J. (2000) Genomic imprinting in plants: observations and evolutionary implications. *Plant Mol. Biol.* **43**, 147–161.

71. Johansson, M.J., He, F., Spatrick, P., Li, C., and Jacobson, A. (2007) Association of yeast Upf1p with direct substrates of the NMD pathway. *Proc. Natl. Acad. Sci. USA* **104**, 20872–20877.

72. Alleman, M., Sidorenko, L., McGinnis, K., Seshadri, V., Dorweiler, J.E., White, J., Sikkink, K., and Chandler, V.L. (2006) An RNA-dependent RNA polymerase is required for paramutation in maize. *Nature* **442**, 295–298.

73. Royce, T.E., Rozowsky, J.S., Bertone, P., Samanta, M., Stolc, V., Weissman, S., Snyder, M., and Gerstein, M. (2005) Issues in the analysis of oligonucleotide tiling microarrays for transcript mapping. *Trends Genet.* **21**, 466–475.

74. Kling, J. (2005) The search for a sequencing thoroughbred. *Nat. Biotechnol.* **23**, 1333–1335.

75. Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M., and Tollervey, D. (1997) The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'–>5' exoribonucleases. *Cell* **91**, 457–466.

76. Yadegari, R., Kinoshita, T., Lotan, O., Cohen, G., Katz, A., Choi, Y., Katz, A., Nakashima, K., Harada, J.J., Goldberg, R.B., Fischer, R.L., and Ohad, N. (2000) Mutations in the FIE and MEA genes that encode interacting polycomb proteins cause parent-of-origin effects on seed development by distinct mechanisms. *Plant Cell* **12**, 2367–2382.