



ELSEVIER



# Mapping the genome landscape using tiling array technology

Junshi Yazaki<sup>1,3</sup>, Brian D Gregory<sup>1,3</sup> and Joseph R Ecker<sup>1,2</sup>

With the availability of complete genome sequences for a growing number of organisms, high-throughput methods for gene annotation and analysis of genome dynamics are needed. The application of whole-genome tiling microarrays for studies of global gene expression is providing a more unbiased view of the transcriptional activity within genomes. For example, this approach has led to the identification and isolation of many novel non-protein-coding RNAs (ncRNAs), which have been suggested to comprise a major component of the transcriptome that have novel functions involved in epigenetic regulation of the genome. Additionally, tiling arrays have been recently applied to the study of histone modifications and methylation of cytosine bases (DNA methylation). Surprisingly, recent studies combining the analysis of gene expression (transcriptome) and DNA methylation (methylome) using whole-genome tiling arrays revealed that DNA methylation regulates the expression levels of many ncRNAs. Further capture and integration of additional types of genome-wide data sets will help to illuminate additional hidden features of the dynamic genomic landscape that are regulated by both genetic and epigenetic pathways in plants.

## Addresses

<sup>1</sup> The Salk Institute for Biological Studies, Plant Molecular and Cellular, Laboratory 10010 North Torrey Pines Road La Jolla, CA 92037, USA

<sup>2</sup> Genomic Analysis Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037, USA

Corresponding author: Ecker, Joseph R. (ecker@salk.edu)

<sup>3</sup> These authors contributed equally to this work.

**Current Opinion in Plant Biology** 2007, **10**:1–9

This review comes from a themed issue on  
Cell Signalling and Gene Regulation  
Edited by Jian-Kang Zhu and Ko Shimamoto

1369-5266/\$ – see front matter

© 2007 Elsevier Ltd. All rights reserved.

DOI 10.1016/j.pbi.2007.07.006

## Introduction

With the completion of several plant genome sequences, and with many more genomes underway, it is now realistic to begin to undertake a variety of experimental and/or computational studies at the whole genome level. One approach that takes advantage of this flood of sequence data to capture genome-wide gene expression information is microarray technology [1–3]. A recent derivation of the standard gene expression microarray is the tiling microarray; these are high-density arrays composed of oligonucleotide probes that span the entire genome of

an organism [4]. The application of numerous experimental procedures can be applied without the requirement of a completely annotated genome [5]. The increase in resolution now provided by tiling arrays will allow dramatic improvements in the understanding of many previously unexplored aspects of genome biology. For instance, using this unbiased approach several plant and animal genomes have been interrogated for (1) alternative splice sites [6–8], (2) transcription unit mapping/genome annotation [9–12], (3) transcription factor binding sites [13–16], (4) comparative genome hybridization [17,18], and (5) for mapping of DNA methylation sites [19,20,21]. In this review, we focus on some recent experiments taking advantage of whole-genome tiling array technology. These studies have revealed that a large portion of the genome previously thought of as being unexpressed actually codes for novel non-protein-coding RNAs (ncRNAs), which have been suggested to be involved in epigenetic regulation. Additionally, tiling arrays have been used to identify a novel epigenetic pathway located within heterochromatic regions (previously viewed as junk DNA) and regulated by DNA methylation [21]. Thus, whole genome tiling microarrays are proving to be a useful tool for functional genomic analyses of a number of biological processes, and more generally for understanding genome complexity.

## Discovery of novel transcription sites using whole-genome tiling microarrays

Although great advances have been made in gene discovery approaches, simply applying computational gene prediction methods for annotation of genome sequences is not sufficient for accurate gene structure determination and/or identification of all transcription units of 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 and non-polyadenylated transcripts. Furthermore, cDNA collections are often devoid of transcripts that are expressed in response to a specific physiological or environmental condition(s). To circumvent such problems, Kapranov *et al.* [22] and Shoemaker *et al.* [23] designed tiling DNA microarrays using sequences homologous to human chromosomes 21 and 22. They used these arrays for gene expression studies by hybridizing targets made from RNA samples extracted from 11 human cell lines. The data obtained from these expression studies identified a large number of novel sites of active gene expression that were not previously annotated by computational gene prediction algorithms using DNA sequencing data or identify in massive collection of

## 2 Cell Signalling and Gene Regulation

sequenced human cDNAs. Additionally, Yamada *et al.* [9] used high-density tiling DNA microarrays containing probes with homology to the entire *Arabidopsis* genome. These arrays were used for gene expression studies by hybridizing targets made from RNA samples of four different tissues (flower, leaf, root, cultured cell). Similar to the expression studies done with human samples, the data obtained from the expression studies in *Arabidopsis* also identified a large number of novel sites of active gene expression in *Arabidopsis* missed by computational gene prediction algorithms and cDNA collections. Interestingly, many of the newly identified transcripts were expressed from the opposite DNA strand in the reverse orientation (anti-sense) to many previously annotated transcripts. Additionally, the study of Yamada *et al.* [9] was able to capture a number of novel transcripts originating in centromeric regions, which were previously thought to be mostly devoid of active gene expression. More recently, Li *et al.* [10] hybridized RNA samples from the indica rice subspecies to high-density tiling microarrays containing probes spanning the entire rice genome. These studies identified a large number of novel sites of active gene expression never before annotated in this model crop genome. Overall, these studies demonstrated that tiling microarrays could be successfully applied for discovery of novel sites of active transcription that lie within the 'dark matter' in genomes.

### Mapping sites of cytosine methylation: the DNA methylome

The methylation of cytosine bases within DNA molecules (DNA methylation) is a heritable epigenetic modification that has been previously demonstrated to regulate the expression of a number of genes without permanent changes to their DNA sequence [24,25]. The regulation of gene expression by DNA methylation can occur in *cis* (the gene itself is methylated) or in *trans* (methylation at another site in the genome regulates the target gene) [26,27]. Recently, whole-genome tiling microarrays were used to map every site of DNA methylation within the *Arabidopsis* genome thus resulting in the first genome-wide high-resolution methylation maps (the methylome), [19<sup>••</sup>, 20<sup>•</sup>]. To do this, these groups used an antibody that specifically recognized methylated cytosine bases within genomic DNA. Therefore, a few hundred base pairs surrounding each region of methylated DNA could be specifically immunoprecipitated using this antibody. The data obtained from these experiments demonstrated that regions containing the highest density of DNA methylation were located in highly repetitive DNA sequences. For instance, centromeric and peri-centromeric repeats of all chromosomes and areas containing large amounts of heterochromatin, like the knob region found on chromosome 4 of *Arabidopsis*, were found to be highly enriched in DNA methylation sites. Interestingly, the genome-wide methylome maps also revealed that over one-third of all previously annotated genes in *Arabidopsis* contain sites of

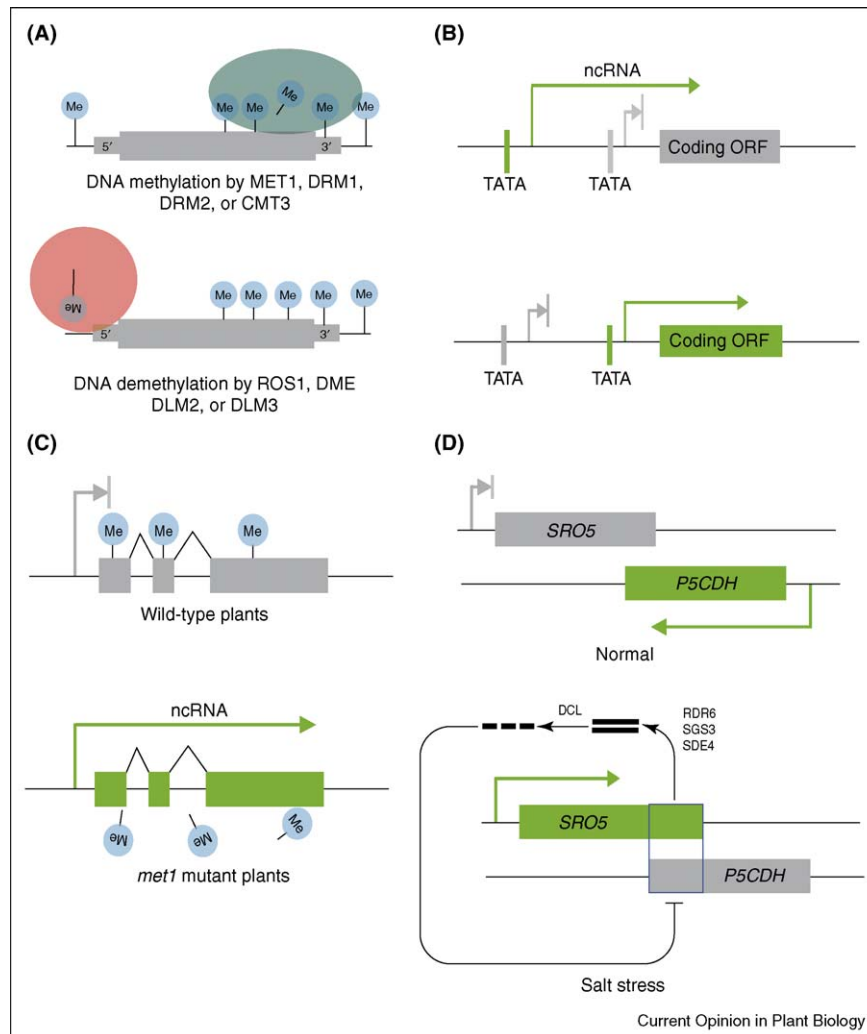
DNA methylation within their transcribed regions, consistent with a previous report [28]. Furthermore, most of the genes that were methylated within their transcribed regions were highly expressed and constitutively active. Additionally, these comprehensive, whole-genome studies demonstrated that the distribution of DNA methylation is clearly different between transposons and genes with annotated function. While DNA methylation of transposons is highly distributed across their entire length including both up and downstream regions, genes with annotated function tended to contain methylation sites with a biased distribution towards their 3' half (Figure 1A). This distribution of DNA methylation may relate to a possible role in silencing of transcription start sites located in the 3' region of open reading frames (ORFs) [29]. Surprisingly, only ~5% of annotated genes contained sites of DNA methylation in regions upstream of their ORFs (promoter regions). In fact, these data suggested that promoter regions are hypomethylated in general. Taken together, such findings suggest that the distribution of DNA methylation may determine how expression from specific genes is effected by this epigenetic mark.

Interestingly, the methyl groups added to cytosine bases within DNA molecules during the process of DNA methylation can also be removed by so called DNA demethylases (Figure 1A). There are four such DNA demethylases in *Arabidopsis*, REPRESSOR OF SILENCING1 (ROS1), DEMETER (DME), DEMETER-LIKE2 (DML2), and DEMETER-LIKE3 (DML3). Interestingly, DME has previously been demonstrated to be required for genomic imprinting during *Arabidopsis* embryo development [30], while the closely related ROS1 is involved in regulating transcriptional gene silencing in a transgenic background mediated by DNA methylation [31]. A recent study by Penterman *et al.* [32<sup>••</sup>] using tiling microarrays to study DNA methylation patterns in wild-type (WT) and mutant plants lacking three of the DNA demethylases (ROS1, DML2, DML3) identified 179 loci that are actively demethylated by DML enzymes in *Arabidopsis*. Therefore, these mutant plants (*ros1-1 dml2-1 dml3-1*) exhibited locus-specific DNA hypermethylation in response to the loss of these three DNA demethylases. Interestingly, demethylation by DML enzymes in gene coding regions primarily occurs at the 5' and 3' ends (Figure 1A), which is a pattern opposite to the overall distribution of WT DNA methylation (see above). Therefore, DNA methylation, which is widely considered a stable epigenetic mark, is actively removed likely to protect genes from potentially deleterious methylation.

### Coupling of the transcriptome and the methylome

Previously, it was determined that DNA methylation and demethylation are critical for controlling the level of gene expression from loci such as *FWA*, *SUPERMAN* [33,34],

Figure 1



Mechanisms of genome regulation revealed by whole genome tiling array. **(A)** A schematic of DNA methylation and demethylation in Arabidopsis. (Top) Whole-genome tiling array revealed that the methyltransferases MET1, DRM1, DRM2, and CMT3 are responsible for the establishment and maintenance of most DNA methylation in Arabidopsis. (Bottom) DML DNA demethylases (ROS1, DML2, and DML3) primarily demethylate the very 5' and 3' ends of open reading frames including promoter and downstream regions. **(B)** Regulation of gene expression by an ncRNA located in the promoter region of an ORF. Transcription of the ncRNA, which is located in the promoter region of an ORF, blocks the binding of transcription factors and/or RNAPII necessary for proper gene expression from this locus. This has been aptly named the transcription interference model for regulation by ncRNAs. **(C)** Expression of intergenic noncoding RNAs regulated DNA methylation. Genome-wide gene expression studies in combination with genome-wide mapping of DNA methylation sites in Arabidopsis identified a number of ncRNAs with increased expression in the absence of DNA methylation. Thus, suggesting that the over-expression of these ncRNAs may indeed be a by-product of the loss of DNA methylation, which when present would act to silence expression from these loci. **(D)** Over-lapping anti-sense gene products regulate *P5CDH* levels during response to salt stress in Arabidopsis. *P5CDH* is expressed under normal Arabidopsis growth conditions. However, salt stress conditions induce expression of *SRO5*, a gene of unknown function. The 3' ends of the *SRO5* and *P5CDH* transcripts overlap, thereby forming a region of dsRNA. This dsRNA region initiates a series of siRNA processing steps that ultimately result in the downregulation of *P5CDH* in response to salt stress. Thus, anti-sense transcripts can regulate neighboring genes through formation of regions of dsRNA that subsequently recruit various RNA silencing pathways. For simplicity only the dsRNA region (blue box) of the two overlapping transcripts is shown in this figure.

*PAI* [35,36], and *BAL* [37]. Additionally, it was suggested that DNA methylation may be involved in suppressing expression of transposons, but experimental evidence for this claim was limited [38–41]. Certain types of transposons are abundant in chromosomal regions near centromeres and constitute a large fraction of the DNA

sequences found in heterochromatic knobs. Interestingly, Lippman *et al.* [38] undertook gene expression studies and methylation mapping using microarrays printed with one-kilobase (kb) PCR products that tiled the entire heterochromatic knob of *Arabidopsis* chromosome 4. They found upon hybridizing target RNA samples

#### 4 Cell Signalling and Gene Regulation

extracted from *met1* mutant plants, which contain a mutation in the gene encoding the maintenance DNA methyltransferase (MET1) required for symmetric (CG) methylation, that transposons and pseudogenes normally not expressed in WT plants were reactivated. Overall, these array studies revealed that in plants containing genetic lesions in *MET1*, gene expression within highly heterochromatic and repetitive sequences was extremely elevated mainly due to the expression of normally silenced transposons and pseudogenes. More recently, Zhang *et al.* [19\*\*] and Zilberman *et al.* [20\*\*] also reported massive transcriptional reactivation of pseudogenes and transposons on an entire genome level, including those found in chromosomal regions rich in heterochromatin and repetitive sequences in *met1* mutant plants using whole-genome tiling arrays.

In addition to a maintenance methyltransferase, plants also have three methyltransferases that are required for *de novo* (non-CG) DNA methylation, DRM1, DRM2, and CMT3. *drm1 drm2 cmt3* triple mutant plants contain genetic lesions in the genes encoding all three *de novo* methyltransferases, and lose a great deal of the non-CG DNA methylation sites in the *Arabidopsis* genome [42]. By hybridizing target RNA samples extracted from *drm1 drm2 cmt3* triple mutant plants to whole-genome tiling arrays, Zhang *et al.* [19\*\*] found that relatively few transposons or pseudogenes were reactivated (in terms of changes in steady state RNA accumulation). In fact, the general transcriptional activity within these triple mutant plants was similar to WT plants at the whole genome level. These findings are consistent with the observation that CG methylation was largely unchanged in heterochromatic genomic regions of triple mutant plants. Furthermore, they suggest that in most cases maintenance (CG) methylation alone is sufficient for the silencing of gene expression from methylated transposons and pseudogenes, and is largely unchanged even in the absence of non-CG methylation. Taken together, the results of these gene expression studies of methyltransferase mutants suggest that the widespread presence of transposons in the centromeres of *Arabidopsis* chromosomes may reflect their importance in heterochromatin formation and/or function [43\*\*].

#### Identification of novel ncRNAs in intergenic regions

One of the most interesting findings from a number of recent high-throughput genomic analyses is the identification and isolation of a large class of previously uncharacterized non-protein-coding RNAs (ncRNAs) (Figure 1B and C). These transcripts were found to be expressed in regions not previously annotated as containing an actively expressed genic unit [9,44]. This class of RNAs seems to represent a major component of the transcriptome that has been suggested to play a novel role in epigenetic gene regulation [25]. Furthermore, a large fraction of these epigenetic regulatory ncRNAs are transcribed by RNA

Polymerase II (RNAPII) [45]. One model for gene silencing mediated by this new class of RNAs suggests that expression of the ncRNA from the same or opposite strand could interfere with expression of a target gene, such as the case for the *SRG1* ncRNA and its target *SER3* (Figure 1B) [46], the *Tsix* ncRNA [47] and its ncRNA target *Xist*, and the *Air* ncRNA [45] and its target *Igf2r*.

Interestingly, the previously mentioned gene expression studies of DNA methyltransferase (*met1* and *drm1 drm2 cmt3*) mutants also identified a large number of ncRNAs whose expression levels increased in these plants [19\*\*]. Therefore, genome-wide gene expression data in combination with mapping of the sites of DNA methylation suggested that the increased expression of these ncRNAs may indeed be a by-product of the loss of DNA methylation, which when present would silence expression from these loci (Figure 1C). Furthermore, a recent report by Kapranov *et al.* [48\*\*] demonstrated that within mammalian genomes there are numerous, small ncRNAs whose expression originates in sites of high-density DNA methylation (CpG islands), suggesting the existence of ncRNAs in mammals that may also be regulated by DNA methylation.

A number of the novel transcripts found in the *Arabidopsis* genome seem to be repeated many of times in the genome (multi-copy elements), thereby suggesting they may represent unannotated transposons [19\*\*]. Furthermore, this class of novel ncRNAs have been suggested to silence neighboring genes through their expression by (1) RNA interference (RNAi) [47,49] and (2) occlusion of promoter binding by RNA polymerase (RNAP) and other transcription factors (Figure 1B) [46,50]. Additionally, others of these novel ncRNAs were found to be single or low copy genic units within the genome [19\*\*]. Many of the ncRNAs in this class have no significant sequence homology to the genomes of other organisms, thus suggesting that the *Arabidopsis* genome contains a large number of fast-evolving ncRNAs whose expression is controlled epigenetically by DNA methylation. Indeed, a very large fraction of ncRNAs are poorly conserved, which has led to the speculation that they are not functional [51]. However, this observation is not necessarily associated to a lack of functionality [52], seeing that the DNA sequences of known functional ncRNAs such as *Xist* and *Air* are poorly conserved [53]. Additionally, an independent study on expression of intergenic sequences in human and chimpanzee showed that conservation of expression of ncRNAs in equivalent genomic positions, but not conservation of the RNA sequences [54]. Therefore, these unique ncRNAs may have originated from evolutionary responses to the growth environment. For example, since plants are sessile organisms, they may have evolved novel genes/processes to rapidly regulate/respond to environmental changes. Interestingly, previous whole-genome gene expression studies using



*Arabidopsis* culture cells by Yamada *et al.* [9] revealed that many pseudogenes, transposons, and ncRNAs are over-expressed in these de-differentiated cells, similar to the transcriptome profile witnessed for *met1* mutant plants. Together, these results suggest that DNA methylation likely plays a role in cellular de-differentiation and regeneration through its regulation of gene expression from putative pseudogenes, transposons, and ncRNAs that litter the *Arabidopsis* genome.

### Identification and regulation of anti-sense RNAs

Massive sequencing efforts have been undertaken to isolate and characterize as many expressed gene products as can possibly be detected from large-scale cDNA libraries. Interestingly, a number of groups involved in these sequencing efforts recognized that there were a far greater number of expressed gene products in their cDNA libraries than potential coding regions predicted by computational methods [44<sup>\*</sup>,55,56]. Upon further examination, these groups found that 70% of the cDNAs that did not fall into potential coding regions, were actually RNAs derived from the strand of DNA opposite to ORFs for computationally predicted protein-coding genes (anti-sense transcripts) [9,57<sup>\*\*</sup>]. More recently, anti-sense transcripts have been found to regulate expression from the ORF on the opposite strand (sense transcript) through formation of a double-stranded RNA (dsRNA) molecule consisting of the sense/anti-sense pair [57<sup>\*\*</sup>]. For instance, Borsani *et al.* [58<sup>\*\*</sup>] reported a natural cis-anti-sense pair of transcripts involved in the regulation of salt tolerance in *Arabidopsis*. This anti-sense gene pair consists of *P5CDH* (encodes  $\Delta^1$ -pyroline-5-carboxylate dehydrogenase) and a gene of unknown function (*SRO5*) that is only expressed upon salt stress conditions. Co-expression of these two genes results in the formation of a dsRNA molecule consisting of the very 3' ends of both transcripts (see Figure 1D). Subsequently, this dsRNA acts as a substrate for small interfering RNA (siRNA) biogenesis. These siRNAs direct initial cleavage of *P5CDH*, thus resulting in a RNA molecule that can act as a substrate for further siRNA biogenesis. Ultimately, this cis-anti-sense pair of transcripts results in the down-regulation of *P5CDH* levels, and improved salt tolerance (Figure 1D). Additionally, with the assistance of whole-genome tiling array data, Swiezewski *et al.* [59<sup>\*</sup>] found that an anti-sense transcript overlapping the 3' end of RNA for the repressor of *Arabidopsis* flowering time, *FLOWERING LOCUS C (FLC)*, is involved in the regulation of flowering time. Furthermore, this group was able to demonstrate that the anti-sense transcript may also act as a biogenesis substrate for siRNAs that are responsible for the heterochromatinization and subsequent silencing of this genomic region. Thus far, dsRNA molecules (sense/anti-sense pairs) have been demonstrated to regulate sense gene expression by (1) inhibition of transcript elongation by RNAP or (2) by acting as a substrate for

siRNA biogenesis, which can result in heterochromatin formation and subsequent silencing of the homologous chromosomal location [57<sup>\*\*</sup>,58<sup>\*\*</sup>,59<sup>\*</sup>].

Previous studies of DNA methylation in a number of organisms have illuminated that a large number of genes are methylated in their coding regions (body-methylated), and suggested this type of methylation can negatively regulate expression from the anti-sense strand [60–62]. Furthermore, the recent combinatorial whole-methylome and transcriptome analyses performed by Zhang *et al.* [19<sup>\*\*</sup>] and Zilberman *et al.* [20<sup>\*</sup>] demonstrated that the majority of constitutively expressed genes are body-methylated (Figure 1A). Zhang *et al.* tested the hypothesis that body-methylation in ORFs negatively regulates anti-sense expression. To do this, the authors analyzed the whole-genome gene expression data for *met1* mutant plants with the hypothesis that if intragenic 'body methylation' silences anti-sense expression in WT plants, then the mutant plants should have increased levels of anti-sense transcripts. However, they found no evidence for a large-scale increase in anti-sense transcription in *met1* mutant plants, which have lost most of the DNA methylation in the body of ORFs, compared to WT plants. These results are consistent with several recent reports concerning the regulation of anti-sense gene expression [63,64]. However, there are possible explanations for why increased anti-sense expression in response to loss of DNA methylation might not be observed. For example, many anti-sense RNAs are likely not to be polyadenylated [65<sup>\*</sup>] and thus, these transcripts would be missed in studies that exclusively use poly-A+ RNA as targets. Additionally, expression of anti-sense transcripts may be regulated by DNA methylation in a tissue or cell-type specific manner. Therefore, understanding of the functions and regulation of DNA methylation and its possible role in anti-sense transcription is at an early stage.

### Genome-wide mapping of histone modifications

In addition to DNA methylation, the covalent addition of methyl groups to lysine or arginine residues of histone tails (histone methylation) has emerged as another crucial step in controlling eukaryotic genome dynamics [66–71]. For instance, tri-methylation of lysine 27 of histone H3 (H3K27me3) plays critical roles in regulating animal development [66–68]. Furthermore, H3K27me3 is required for proper regulation of several genes important for development in plants [69–71]. Recently, genome-wide profiling of H3K27me3 in the *Arabidopsis* genome was carried out using tiling microarrays [72<sup>\*\*</sup>]. Interestingly, the results from this study suggest that H3K27me3 is a major silencing mechanism in plants that regulates an unexpectedly large number of *Arabidopsis* genes. Furthermore, analysis of the whole-genome profile of H3K27me3 suggested that establishment and maintenance of this

## 6 Cell Signalling and Gene Regulation

specific epigenetic modification is largely independent of other epigenetic pathways, such as DNA methylation or RNA silencing. Therefore, the use of whole-genome tiling microarrays to study a variety of epigenetic regulatory pathways in plants and animals have suggested an extremely complex network of mechanisms involved in regulation of genome dynamics. Furthermore, through the combination these various data types only now can we begin to gain an appreciation for the complex nature of regulatory mechanisms employed in controlling genome dynamics.

### Conclusions and future outlook

The use of unbiased genome-wide approaches for characterization of the genomic landscape in numerous organisms is well underway. For instance, in *Arabidopsis* alone high-throughput analysis of the transcriptome, the DNA and histone methylomes, and more recently, for comparisons of diversity among various ecotypes [73,74] using microarrays are now providing reams of genome data for this reference plant. In addition to microarrays, new technologies for extremely deep DNA sequencing (e.g. Solexa/Illumina, ABI SOLID, 454/Roche, etc.) are also rapidly emerging. For instance, a recent report by Barski *et al.* [75] demonstrates high-resolution profiling of 20 histone modifications and protein–DNA interactions in the human genome at single base resolution using Solexa 1G sequencing.

Ultra high-throughput DNA sequencing can be viewed as a ‘virtual’ tiling array, where hundreds of millions of short read sequences will push genomic analysis to an even higher level of resolution [75,76]. Unlike microarrays, next-generation DNA sequencing technologies are not limited to sequenced genomes. Analysis of short-nucleotide-fragments like transcription factor binding sites [75,76] and small RNAs (smRNAs), can result in direct identification of the sequences of these genomic targets. Additionally, next-generation sequencing can also be employed for whole transcriptome sequencing and for large-scale genome (re)sequencing for the detection of genetic variation amongst various individuals within a population. Interestingly, a recent study by Euskirchen *et al.* [77] compared strategies for mapping transcription factor (TF) binding sites in the human genome using chromatin immuno-precipitation (ChIP) in combination with two different whole-genome methodologies, DNA microarray analysis (ChIP-chip) and DNA sequencing (ChIP-PET). Comparison of these two methods revealed strong agreement for the highest ranked binding sites with less overlap for the lowest ranked targets. With advantages and disadvantages unique to each approach, this study illuminated the idea that ChIP-chip and ChIP-PET are frequently complementary in their relative abilities to detect transcription factor binding sites targets, whereby the most comprehensive and complete list of binding regions is obtained by merging results from

ChIP-chip and ChIP-sequencing. Therefore, whatever the technology to be employed, these genome-wide approaches have already begun to change the scope of our understanding of genome dynamics in only a few short years. Their continued development and implementation for capturing a deeper spectrum of genome-wide data will provide for an even greater understanding of the complexity of genetic and epigenetic control mechanisms.

### Acknowledgements

Work in our laboratory is supported by grants to J.R.E. from the National Science Foundation, the Department of Energy and the National Institutes of Health. B.D.G is a Damon Runyon Fellow supported by the Damon Runyon Cancer Research Foundation.

### References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
  - of outstanding interest
1. Schena M, Shalon D, Davis RW, Brown PO: **Quantitative monitoring of gene expression patterns with a complementary DNA microarray.** *Science* 1995, **270**:467-470.
  2. Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW: **Parallel human genome analysis: microarray-based expression monitoring of 1000 genes.** *Proc Natl Acad Sci U S A* 1996, **93**:10614-10619.
  3. Lashkari DA, DeRisi JL, McCusker JH, Namath AF, Gentile C, Hwang SY, Brown PO, Davis RW: **Yeast microarrays for genome wide parallel genetic and gene expression analysis.** *Proc Natl Acad Sci U S A* 1997, **94**:13057-13062.
  4. Borevitz JO, Ecker JR: **Plant genomics: the third wave.** *Annu Rev Genomics Hum Genet* 2004, **5**:443-477.
  5. Mockler TC, Ecker JR: **Applications of DNA tiling arrays for whole-genome analysis.** *Genomics* 2005, **85**:1-15.
  6. Johnson JM, Castle J, Garrett-Engle P, Kan Z, Loerch PM, Armour CD, Santos R, Schadt EE, Stoughton R, Shoemaker DD: **Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays.** *Science* 2003, **302**:2141-2144.
  7. Clark TA, Sugnet CW, Ares M Jr: **Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays.** *Science* 2002, **296**:907-910.
  8. Castle J, Garrett-Engle P, Armour C, Duenwald S, Loerch P, Meyer M, Schadt E, Stoughton R, Parrish M, Shoemaker D *et al.*: **Optimization of oligonucleotide arrays and RNA amplification protocols for analysis of transcript structure and alternative splicing.** *Genome Biol* 2003, **4**:R66.
  9. Yamada K, Lim J, Dale JM, Chen H, Shinn P, Palm CJ, Southwick AM, Wu HC, Kim C, Nguyen M *et al.*: **Empirical analysis of transcriptional activity in the *Arabidopsis* genome.** *Science* 2003, **302**:842-846.
  10. Li L, Wang X, Stolc V, Li X, Zhang D, Su N, Tongprasit W, Li S, Cheng Z, Wang J *et al.*: **Genome-wide transcription analyses in rice using tiling microarrays.** *Nat Genet* 2006, **38**:124-129.
  11. Rozowsky J, Wu J, Lian Z, Nagalakshmi U, Korbel JO, Kapranov P, Zheng D, Dyke S, Newburger P, Miller P *et al.*: **Novel transcribed regions in the human genome.** *Cold Spring Harb Symp Quant Biol* 2006, **71**:111-116.
  12. Stolc V, Samanta MP, Tongprasit W, Sethi H, Liang S, Nelson DC, Hegeman A, Nelson C, Rancour D, Bednarek S *et al.*: **Identification of transcribed sequences in *Arabidopsis thaliana* by using high-resolution genome tiling arrays.** *Proc Natl Acad Sci U S A* 2005, **102**:4453-4458.

13. Thibaud-Nissen F, Wu H, Richmond T, Redman JC, Johnson C, Green R, Arias J, Town CD: **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* 2006, **47**:152-162.
14. Katou Y, Kanoh Y, Bando M, Noguchi H, Tanaka H, Ashikari T, Sugimoto K, Shirahige K: **S-phase checkpoint proteins Tof1 and Mrc1 form a stable replication-pausing complex.** *Nature* 2003, **424**:1078-1083.
15. Taverner N, Smith J, Wardle F: **Identifying transcriptional targets.** *Genome Biol* 2004, **5**:210.
16. Buck MJ, Lieb JD: **ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments.** *Genomics* 2004, **83**:349-360.
17. Ishkanian AS, Malloff CA, Watson SK, deLeeuw RJ, Chi B, Coe BP, Snijders A, Albertson DG, Pinkel D, Marra MA *et al.*: **A tiling resolution DNA microarray with complete coverage of the human genome.** *Nat Genet* 2004, **36**:299-303.
18. Bignell GR, Huang J, Greshock J, Watt S, Butler A, West S, Grigoroava M, Jones KW, Wei W, Stratton MR *et al.*: **High-resolution analysis of DNA copy number using oligonucleotide microarrays.** *Genome Res* 2004, **14**:287-295.
19. Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SWL, Chen H, ● Henderson IR, Shinn P, Pellegrini M, Jacobsen SE *et al.*: **Genome-wide high-resolution mapping and functional analysis of DNA methylation in Arabidopsis.** *Cell* 2006, **126**:1189-1201.
- This is the first combinatorial study of the DNA methylome and transcriptome using whole-genome tiling microarrays in any organism.
20. Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S: ● **Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription.** *Nat Genet* 2007, **39**:61-69.
- This paper reports whole DNA methylome and transcriptome maps for Arabidopsis. These studies demonstrated that gene expression and DNA methylation are closely interwoven processes.
21. Martienssen R, Doerge R, Colot V: **Epigenomic mapping in Arabidopsis using tiling microarrays.** *Chromosome Res* 2005, **13**:299-308.
22. Kapranov P, Cawley SE, Drenkow J, Bekiranov S, Strausberg RL, Fodor SPA, Gingeras TR: **Large-scale transcriptional activity in chromosomes 21 and 22.** *Science* 2002, **296**:916-919.
23. Shoemaker DD, Schadt EE, Armour CD, He YD, Garrett-Engle P, McDonagh PD, Loerch PM, Leonardson A, Lum PY, Cavet G *et al.*: **Experimental annotation of the human genome using microarray technology.** *Nature* 2001, **409**:922-927.
24. Chan SWL, Henderson IR, Jacobsen SE: **Gardening the genome: DNA methylation in Arabidopsis thaliana.** *Nat Rev Genet* 2005, **6**:351-360.
25. Schaefer CB, Ooi SKT, Bestor TH, Bourc'his D: **Epigenetic decisions in mammalian germ cells.** *Science* 2007, **316**:398-399.
26. Chandler VL: **Poetry of b1 Paramutation: cis- and trans-chromatin communication.** *Cold Spring Harb Symp Quant Biol* 2004, **69**:355-362.
27. Chandler VL, Stam M: **Chromatin conversations: mechanisms and implications of paramutation.** *Nat Rev Genet* 2004, **5**:532-544.
28. Tran RK, Henikoff JG, Zilberman D, Ditt RF, Jacobsen SE, Henikoff S: **DNA methylation profiling identifies CG methylation clusters in Arabidopsis genes.** *Current Biol* 2005, **15**:154-159.
29. Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Sempke CAM, Taylor MS, Engstrom PG, Frith MC *et al.*: **Genome-wide analysis of mammalian promoter architecture and evolution.** *Nat Genet* 2006, **38**:626-635.
30. Choi Y, Gehring M, Johnson L, Hannon M, Harada JJ, Goldberg RB, Jacobsen SE, Fischer RL: **DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in Arabidopsis.** *Cell* 2002, **110**:33-42.
31. Zhu J, Kapoor A, Sridhar VV, Agius F, Zhu JK: **The DNA glycosylase/lyase ROS1 functions in pruning DNA methylation patterns in Arabidopsis.** *Current Biol* 2007, **17**:54-59.
32. Penterman J, Zilberman D, Huh JH, Ballinger T, Henikoff S, ● Fischer RL: **DNA demethylation in the Arabidopsis genome.** *Proc Natl Acad Sci U S A* 2007, **104**:6752-6757.
- The authors identified regions of active demethylation mediated by DNA glycosylases throughout the entire Arabidopsis genome. Additionally, they also assert that demethylation by DNA glycosylases is required to edit the pattern of DNA methylation within genomes to protect genes from potentially deleterious methylation.
33. Jacobsen SE, Meyerowitz EM: **Hypermethylated SUPERMAN epigenetic alleles in Arabidopsis.** *Science* 1997, **277**:1100-1103.
34. Soppe WJJ, Jacobsen SE, Alonso-Blanco C, Jackson JP, Kakutani T, Koornneef M, Peeters AJM: **The late flowering phenotype of fwa mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene.** *Mol Cell* 2000, **6**:791-802.
35. Bender J, Fink GR: **Epigenetic control of an endogenous gene family is revealed by a novel blue fluorescent mutant of Arabidopsis.** *Cell* 1995, **83**:725-734.
36. Melquist S, Luff B, Bender J: **Arabidopsis PAI gene arrangements, cytosine methylation and expression.** *Genetics* 1999, **153**:401-413.
37. Stokes TL, Kunkel BN, Richards EJ: **Epigenetic variation in Arabidopsis disease resistance.** *Genes Dev* 2002, **16**:171-182.
38. Lippman Z, Martienssen R: **The role of RNA interference in heterochromatic silencing.** *Nature* 2004, **431**:364-370.
39. Miura A, Yonebayashi S, Watanabe K, Toyama T, Shimada H, Kakutani T: **Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis.** *Nature* 2001, **411**:212-214.
40. Hirochika H, Okamoto H, Kakutani T: **Silencing of retrotransposons in Arabidopsis and reactivation by the ddm1 mutation.** *Plant Cell* 2000, **12**:357-369.
41. Singer T, Yordan C, Martienssen RA: **Robertson's mutator transposons in A. thaliana are regulated by the chromatin-remodeling gene decrease in DNA methylation (DDM1).** *Genes Dev* 2001, **15**:591-602.
42. Chan SWL, Henderson IR, Zhang X, Shah G, Chien JSC, Jacobsen SE: **RNAi, DRD1, and histone methylation actively target developmentally important non-CG DNA methylation in Arabidopsis.** *PLoS Genet* 2006, **2**:e83.
43. Zaratiegui M, Irvine DV, Martienssen RA: **Noncoding RNAs and ● gene silencing.** *Cell* 2007, **128**:763-776.
- A comprehensive review of non-protein-coding transcripts (ncRNAs) and gene silencing in various organisms. The authors review the role of ncRNA in heterochromatic silencing, the silencing of transposable elements. Furthermore, they discussed the role of co-transcriptional processing by RNA interference and other mechanisms, as well as parallels of RNA silencing with the processes of imprinting, paramutation, polycomb silencing, and X inactivation.
44. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, ● Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C *et al.*: **The FANTOM Consortium: The transcriptional landscape of the mammalian genome.** *Science* 2005, **309**:1559-1563.
- The authors provide a comprehensive description of transcription start and termination sites using a massive full-length cDNA collection.
45. Seidl CI, Stricker SH, Barlow DP: **The imprinted Air ncRNA is an atypical RNAPII transcript that evades splicing and escapes nuclear export.** *EMBO J* 2006, **25**:3565-3575.
46. Martens JA, Laprade L, Winston F: **Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene.** *Nature* 2004, **429**:571-574.



## 8 Cell Signalling and Gene Regulation

47. Sado T, Hoki Y, Sasaki H: **Tsix defective in splicing is competent to establish Xist silencing.** *Development* 2006, **133**:4925-4931.
48. Kapranov P, Cheng J, Dike S, Nix DA, Dutttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermueller J, Hofacker IL *et al.*: **RNA maps reveal new RNA classes and a possible function for pervasive transcription.** *Science* 2007:1138341.
- This study investigated the genomic origins and relationships of human nuclear and cytosolic polyadenylated RNAs both longer and shorter than 200 nucleotides in length. From their results, the authors suggest a role for some unannotated RNAs as primary transcripts for the production of short RNAs.
49. Andersen AA, Panning B: **Epigenetic gene regulation by noncoding RNAs.** *Curr Opin Cell Biol* 2003, **15**:281-289.
50. Hirschman JE, Durbin KJ, Winston F: **Genetic evidence for promoter competition in *Saccharomyces cerevisiae*.** *Mol Cell Biol* 1988, **8**:4608-4615.
51. Babak T, Blencowe B, Hughes T: **A systematic search for new mammalian noncoding RNAs indicates little conserved intergenic transcription.** *BMC Genom* 2005, **6**:104.
52. Carninci P, Hayashizaki Y: **Noncoding RNA transcription beyond annotated genes.** *Curr Opin Genet Dev* 2007, **17**:139-144.
53. Pang KC, Frith MC, Mattick JS: **Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function.** *Trends Genet* 2006, **22**:1-5.
54. Khaitovich P, Kelso J, Franz H, Visagie J, Giger T, Joerchel S, Petzold E, Green RE, Lachmann M, Pääbo S: **Functionality of intergenic transcription: an evolutionary comparison.** *PLoS Genet* 2006, **2**:e171.
55. Suzuki M, Hayashizaki Y: **Mouse-centric comparative transcriptomics of protein coding and non-coding RNAs.** *Bioessays* 2004, **26**:833-843.
56. Claverie JM: **Fewer genes, more noncoding RNA.** *Science* 2005, **309**:1529-1530.
57. Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, Nakamura M, Nishida H, Yap CC, Suzuki M, Kawai J *et al.*: RIKEN Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group) and the FANTOM Consortium: **Antisense transcription in the mammalian transcriptome.** *Science* 2005, **309**:1564-1566.
- The authors demonstrate a global analysis of the mouse transcriptome, which provides evidence that a large proportion of the genome can produce transcripts from both strands (anti-sense gene pairs). Additionally, the researchers find that anti-sense transcripts commonly join neighboring 'genes' into complex loci of linked transcription.
58. Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu JK: **Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in *Arabidopsis*.** *Cell* 2005, **123**:1279-1291.
- This paper describes an over-lapping anti-sense gene pair that generates a region of dsRNA. This dsRNA region initiates a series of siRNA processing steps that ultimately result in the downregulation of *P5CDH* in response to salt stress. From their results, the authors suggest that anti-sense transcripts can regulate neighboring genes through formation of regions of dsRNA that subsequently recruit various RNA silencing pathways.
59. Swiezewski S, Crevillen P, Liu F, Ecker JR, Jerzmanowski A, Dean C: **Small RNA-mediated chromatin silencing directed to the 3' region of the *Arabidopsis* gene encoding the developmental regulator, FLC.** *Proc Natl Acad Sci U S A* 2007, **104**:3633-3638.
- The authors report the identification of two smRNAs corresponding to the reverse strand 3' to the canonical poly(A) site of the gene encoding a repressor of flowering in *Arabidopsis*, *FLOWERING LOCUS C* (FLC), using whole genome tiling array data. These smRNAs are required for proper regulation of FLC expression and ultimately flowering time.
60. Hohn T, Corsten S, Rieke S, Muller M, Rothnie H: **Methylation of coding region alone inhibits gene expression in plant protoplasts.** *Proc Natl Acad Sci U S A* 1996, **93**:8334-8339.
61. Keshet I, Yisraeli J, Cedar H: **Effect of regional DNA methylation on gene expression.** *Proc Natl Acad Sci U S A* 1985, **82**:2560-2564.
62. Barry C, Faugeron G, Rossignol J: **Methylation induced premeiotically in ascobolus: coextension with DNA repeat lengths and effect on transcript elongation.** *Proc Natl Acad Sci U S A* 1993, **90**:4557-4561.
63. Faghihi M, Wahlestedt C: **RNA interference is not involved in natural antisense mediated regulation of gene expression in mammals.** *Genome Biol* 2006, **7**:R38.
64. Wang XJ, Gaasterland T, Chua NH: **Genome-wide prediction and identification of cis-natural antisense transcripts in *Arabidopsis thaliana*.** *Genome Biol* 2005, **6**:R30.
65. Gehring M, Henikoff S: **DNA methylation dynamics in plant genomes.** *Biochim Biophys Acta*, In press.
- A comprehensive review on the functions of DNA methylation in plants. The authors review the impact of DNA methylation on transposable elements, genes, and on the genome as a whole. Additionally, the extent to which changes in DNA methylation are part of the plant life cycle are also discussed.
66. Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, Chevalier B, Johnstone SE, Cole MF, Isono KI *et al.*: **Control of developmental regulators by polycomb in human embryonic stem cells.** *Cell* 2006, **125**:301-313.
67. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K *et al.*: **A bivalent chromatin structure marks key developmental genes in embryonic stem cells.** *Cell* 2006, **125**:315-326.
68. Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI, Levine SS, Wernig M, Tajonar A, Ray MK *et al.*: **Polycomb complexes repress developmental regulators in murine embryonic stem cells.** *Nature* 2006, **441**:349-353.
69. Schubert D, Clarenz O, Goodrich J: **Epigenetic control of plant development by Polycomb-group proteins.** *Curr Opin Plant Biol* 2005, **8**:553-561.
70. Kinoshita T, Harada JJ, Goldberg RB, Fischer RL: **Polycomb repression of flowering during early plant development.** *Proc Natl Acad Sci U S A* 2001, **98**:14156-14161.
71. Chanvattana Y, Bishopp A, Schubert D, Stock C, Moon YH, Sung ZR, Goodrich J: **Interaction of Polycomb-group proteins controlling flowering in *Arabidopsis*.** *Development* 2004, **131**:5263-5276.
72. Zhang X, Clarenz O, Cokus S, Bernatavichute YV, Pellegrini M, Goodrich J, Jacobsen SE: **Whole-genome analysis of histone H3 lysine 27 trimethylation in *Arabidopsis*.** *PLoS Biol* 2007, **5**:e129.
- The authors report identification of regions containing tri-methylation of histone H3 lysine 27 (H3K27me3) in the *Arabidopsis* genome using whole genome tiling arrays. They suggest that H3K27me3 is a major silencing mechanism in plants that regulates an unexpectedly large number of genes, and this silencing pathway is unrelated to others that have been previously identified.
73. Borevitz JO, Michael TP, Hazen SP, Morris GP, Baxter IR, Hu TT, Chen H, Werner J, Salt DE, Kay SA *et al.*: **Proc Natl Acad Sci U S A** 2007, **104**:12057-12062.
- Genome-wide patterns of single feature polymorphism in *Arabidopsis thaliana*.
74. Clark RM, Schweikert G, Toomajian C, Ossowski S, Zeller G, Shinn P, Warthmann N, Hu TT, Fu G, Hinds DA *et al.*: **Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*.** *Science* 2007, **317**:338-342.
75. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K: **High-resolution profiling of histone methylations in the human genome.** *Cell* 2007, **129**:823-837.
- This is the first, comprehensive, whole-genome mapping of numerous histone modifications as well as binding sites for histone variant H2A.Z, RNA polymerase II, and the insulator binding protein CTCF in any organism using ultra high-throughput Solexa 1G sequencing technology.
76. Johnson DS, Mortazavi A, Myers RM, Wold B: **Genome-wide mapping of *in vivo* protein-DNA interactions.** *Science* 2007, **316**:1497-1502.



The authors develop a large-scale chromatin immunoprecipitation assay (ChIP-Seq) for the transcription factor neuron-restrictive silencer factor (NRSF; also known as REST, for repressor element-1 silencing transcription factor), and then utilized ultra-high-throughput DNA Solexa 1G sequencing technology to map a genome-wide binding profile for this transcription factor. They find that NRSF binds to 1946 locations within the human genome, including a number of novel genomic locations. Additionally, they use this high-density data to map an extremely

accurate consensus recognition sequence for this very important transcription factor.

77. Euskirchen GM, Rozowsky JS, Wei CL, Lee WH, Zhang ZD, Hartman S, Emanuelsson O, Stolc V, Weissman S, Gerstein MB *et al.*: **Mapping of transcription factor binding regions in mammalian cells by ChIP: Comparison of array- and sequencing-based technologies.** *Genome Res* 2007, **17**:898-909.