GENOMICS

Mapping the plant methylome

Bisulphite sequencing of the *Arabidopsis* genome using Solexa technology yields a map of methylated cytosines at single-base-pair resolution.

The four nucleotides that make up the primary sequence of DNA are not the sole determinants of gene expression: there is a fifth player with a prominent role—methylcytosine, also referred to as the fifth base. Addition of a methyl group to cytosine is associated with stable and heritable repression of transcription.

Finding genes and genomic regions that are silenced during development or disease processes is of great interest to researchers; methylcytosines serve as markers, but precisely locating the distribution of these markers in the whole genome is a challenge.

It was such a fine-scale map of genome-wide methylation sites that Steve Jacobsen and computational biologist Matteo Pellegrini from the University of California at Los Angeles (UCLA) wanted to develop. Their team chose the relatively small genome of the plant *Arabidopsis thaliana* to develop and hone the necessary tools and techniques (**Fig. 1**).

Over the years researchers have used DNA microarrays to locate the methylcytosines, but some of the shortcomings of microarrays left Jacobsen and Pellegrini looking for new techniques. Jacobsen explains: "One of the limitations of arrays is that they are not single-base-resolution," and single-base resolution was precisely what the scientists needed to analyze the methylation status of every cytosine in the *Arabidopsis* genome; so they turned to high-throughput sequencing.

To prepare the plant DNA for sequencing, the scientists had two requirements: they needed to distinguish methylated cytosines from the non-methylated versions by their sequence and break the genome into short fragments flanked by the sequencing primers.

Jacobsen's team combined the construction of the short sequence DNA library with the well-established technique of bisulphite conversion, during which unmethylated cytosines are converted to uracil and ultimately thymidine, while the methylated cytosines remain unchanged. To ensure complete conversion of cytosines, the researchers ligated primers to their randomly sheared DNA before the bisulphite treatment, and then used a second set of primers—which only anneal to the first set if complete conversion has taken place—to amplify the library. This second set or primers contained a 5-nucle-

Figure 1 | *Arabidopsis thaliana*.

GENE REGULATION

INSULATING GENE EXPRESSION

In an analysis of transgene insertion sites, researchers find that position effects cause tissue-specific expression of transgenes but that insulators allow uniform expression.

Transgenes are subject to endogenous regulatory regions, manifesting themselves as position effects that enhance or, more typically, reduce transgene expression. Now, postdoc Michele Markstein and her colleagues in Norbert Perrimon's lab at Harvard University report a systematic comparison of transgene expression loci in *Drosophila melanogaster*. They found that transgene expression varies between tissues because of position effects at the insertion site. But the addition of insulator elements, which shield the intervening DNA from position effects and probably affect the transgene's chromatin modifications as well, canceled these position effects and boosted transgene expression ubiquitously.

For *Drosophila*, in which homologous recombination is painfully inefficient, protocols for the randomized insertion of transgenes are giving way to site-specific systems. But, as Markstein says, "targeting transgenes to a single locus does not by itself solve the problem of position effects because all loci have position effects. The question remains, 'which locus would be ideal to target?'." The answer, it turns out, depends on what you are trying to do. Many high-throughput studies need to maximize expression to generate an easily scorable phenotype. For instance, based on earlier experiments, the Harvard group knew that "to create an efficient [RNA interference] library we'd need to find a way to ensure that the transgenes were expressed at sufficiently high levels to produce a phenotype," says Markstein. So Markstein and her colleagues set out to find the 'golden locus' for efficient insertion of highly expressing transgenes.

They used a virus–derived integration system, for which a plasmid that contains an insertion element lands in a landing element in the genome. With this system, Markstein and her colleagues targeted luciferase-containing transgenes to allow them to quantitatively measure expression at each of 20 landing sites in the genome. When the researchers induced transgene expression ubiquitously, they detected differing amounts of luciferase with each fly line, reflecting varying position effects at the transgene loci, as expected.

When the researchers expressed the transgene in different tissues, however, they saw that the position effects at each transgene locus and the resulting expression of the transgenes varied greatly between tissues. In one particularly striking example, two loci that had similar expression in the nervous system had a 20-fold difference in muscle. These observations underscore previous studies, which found that so-called ubiquitous expression loci have tissue-by-tissue variability. For instance, Markstein points out that the popular mouse *Gt(ROSA)26Sor* (also known as *ROSA26*)

otide tag that would allow them to orient the read and determine whether it came from the sense or the antisense strand.

To efficiently deal with the enormous amount of data generated by high-throughput sequencing technology, Jacobsen and Pellegrini developed algorithms that improved the quality of the base called during the actual sequencing procedure, allowed better mapping of the reads to the genome and filtered out any reads that still contained unconverted cytosines.

They installed filters in their analysis program that eliminated all reads that did not uniquely map to the genome and ended up with a DNA methylation map that comprised 84% of the plant genome.

The results speak to the increased sensitivity of this bisulphite-sequencing method over microarray-based techniques. The team at UCLA was able to find new methylation sites in genes previously classified as unmethylated; they mapped methylation across highly repetitive ribosomal DNA loci and accurately detected methylated promoters.

Of course such high-resolution methylation mapping is of interest not only to the plant community. Jacobsen is certain that their approach is transferable to higher organisms such as mouse and human. He sees the main limitation at this point in the high cost of sequencing for large genomes but adds confidently: "sequencing technologies are improving their throughput at a fast pace, so this technique will be practical quickly."

Detailed methylation patterns may soon be as self evident a resource as primary genomic sequences are at the moment.

Nicole Rusk

RESEARCH PAPERS

Cokus, S.J. *et al.* Shotgun bisulphate sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* **452**, 215–219 (2008).

locus "is inefficient for many studies because it does not drive uniformly high levels of expression in all tissues."

But how to counter these tissue-specific position effects to induce truly ubiquitous expression? The researchers flanked the transgenes with insulator elements. Addition of these insulators profoundly increased expression of the transgenes in all tissues and effectively blocked the tissue-specific position effects. So although the researchers did not find a 'golden locus' that could produce uniform, ubiquitous expression, they could create the same effect with insulator elements while making transgenic flies efficiently with site-specific integration. This particular insulator acts similarly to vertebrate CTCF insulators, suggesting the latter might also boost transgene expression in mice. But for flies, researchers at Janelia Farm are already using this strategy to develop a long hoped-for RNAi library.

These transgene landing site–containing fly lines have other uses as well. As Markstein proposes, "a series of sites with different levels of inducibility in a particular tissue may be targeted to create a controlled allelic series, or a single site may be selected because of its high or low inducibility in a specific tissue." Ultimately, though, she hopes to use these fly lines to study what makes them so capricious in the first place: position effects.

Katherine Stevens

RESEARCH PAPERS

Markstein, M. *et al.* Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. *Nat. Genet.*, published online 2 March 2008.

NEWS IN BRIEF

MICROSCOPY

High-speed super-resolution imaging

Methods for imaging fluorescent samples at resolutions much greater than possible with conventional imaging have only begun to be applied to living cells. Westphal *et al.* adapted one of the earliest super-resolution methods, stimulated emission depletion (STED) microscopy, for video-rate super-resolution imaging of fluorescently labeled synaptic vesicles in living cells. These structures were ideally suited for a first demonstration of this method but improvements should permit application to other systems.

Westphal, V. *et al. Science*, published online 21 February 2008.

PROTEOMICS

Cracking the histone H4 code

The post-translational modifications on histone tails known as 'codes' guide DNA-chromatin interactions. Phanstiel *et al.* describe a method using nanoflow high-performance liquid chromatography to separate intact histone tails, combined with high-resolution mass spectrometry–based sequencing, and applied it to decipher the combinatorial histone H4 codes in human embryonic stem cells undergoing differentiation.

Phanstiel, D. *et al. Proc. Natl. Acad. Sci. USA*, published online 7 March 2008.

IMAGING AND VISUALIZATION

Lighting up synapses

Assessing synaptic connectivity in the dense nerve bundles of the nervous system is very challenging. Feinberg *et al.* describe a method to label synapses in *Caenorhabditis elegans* by expressing complementary GFP fragments tethered to transmembrane proteins on different cells. With complementation of the GFP fragments, fluorescence is restored, and this signals the proximity of the presynaptic and postsynaptic plasma membranes.

Feinberg, E.H. *et al. Neuron* **57**, 353–363 (2008).

PROTEIN BIOCHEMISTRY

Counting disulfide bonds

Large-scale structural analysis of proteins containing multiple disulfide bonds has been difficult owing to the absence of methods for distinguishing their native forms from misfolded intermediates. Narayan *et al.* now describe a method that uses mild reduction to selectively reduce the less stable non-native disulfide bonds and chemical blocking of free cysteines, coupled with mass spectrometry to determine the number of disulfide bonds, thus allowing native forms of proteins in mixtures to be distinguished.

Narayan, M. *et al. Nat. Biotechnol.*, published online 17 February 2008.

BIOPHYSICS

Molecular cutting and pasting

Kufer *et al.* describe a method for assembling biomolecular structures in defined geometric patterns using atomic force microscopy. By taking advantage of the natural ability of DNA to hybridize and by applying different unbinding forces that act on different DNA geometries, they show that target molecules coupled to DNA oligomers can be picked from one area on a surface with an AFM tip, moved and deposited in a new location.

Kufer, S.K. *et al. Science* **319**, 594–596 (2008).