Minireview

Pyrosequencing positions nucleosomes precisely
Tom Owen-Hughes and Maik Engeholm

Address: Division of Gene Regulation and Expression, College of Life Sciences, University of Dundee, Dundee DD1 5EH, UK.

Correspondence: Tom Owen-Hughes. Email: t.a.owenhughes@dundee.ac.uk

Abstract

New parallel-sequencing technology has recently been used to map with unprecedented accuracy the positions of nucleosomes enriched for the histone variant H2A.Z throughout the yeast genome.

Determining where nucleosomes are positioned across eukaryotic genomes, and why, has long been one of the major aims of chromatin research. A recent study in *Nature* [1] has advanced the field greatly by providing the most accurate map of nucleosomes to date. Historically, a range of different strategies has been used to map their positions. Many relied on the ability of nucleosomes to protect DNA from digestion by nucleases. Indeed, digestion with micro-coccal nuclease (MNase) was responsible for originally defining the nucleosome as containing 146 bp of DNA [2]. Combining nuclease digestion with Southern blotting proved a powerful way of mapping regions that are hypersensitive to nuclease digestion. Improvements to the resolution of this approach eventually enabled the positions of individual nucleosomes to be mapped [3]. Nucleosome positions could be assigned with even greater accuracy by using the nucleosomal DNA as a substrate for primer extension [4].

More recently, it has become possible to assess nucleo-some positioning over large regions using high-density tiled microarrays to identify the undigested protected DNA. This approach has enabled some general features of the nucleosomal landscape to be identified; for example, in budding yeast the region in the vicinity of the transcription start site is depleted of nucleosomes [5,6]. A limitation of this approach has been that the spacing between oligonucleotides on the arrays and the length of the oligonucleotides themselves limit the accuracy with which nucleosome positions can be assigned. All these approaches rely on characterizing chromatin following

treatment of the cells to make them permeable to nucleases, with the risk that alterations to chromatin structure could occur during this procedure. In an attempt to avoid this, a range of strategies has been used to map chromatin structure in living cells [7].

An alternative approach is to use bioinformatics to predict nucleosome positions on a genome-wide scale. On the assumption that nucleosome positions are, at least in part, determined by the interactions with the underlying DNA, a relatively small collection of nucleosomes can be used to build a model of nucleotide sequences preferentially occupied by nucleosomes. This approach has the potential to assign nucleosome positions with 1-bp resolution. It has proved particularly informative to look at the distribution of individual dinucleotide steps around the nucleosome. It turns out that in nucleosomes isolated both from native chromatin and from *in vitro* reconstitution assays, characteristic dinucleotide steps occur with an approximately 10-bp periodicity at specific positions of the nucleosome core particle (NCP) DNA [8,9]. On the basis of such a model it has been possible to predict nucleosome organization over the entire yeast genome, and the predictions in many instances appear to coincide with experimentally determined positions. In certain regions of the genome, however, nucleosomes are actively being displaced from sequences with the most favorable nucleosome-DNA interactions [10]. Such changes cannot be predicted by current models, meaning that there is still a requirement to measure nucleosome occupancy experimentally.

Pyrosequencing to the rescue

In a paper recently published in *Nature*, Albert *et al.* [1] have overcome the loss of resolution attributable to the length and spacing of oligonucleotides on arrays by directly sequencing the products of a MNase digestion of *Saccharomyces cerevisiae* chromatin, which enables the ends of each MNase-trimmed fragment to be identified with base-pair precision. This approach also involves sequencing very large numbers of nucleosomal fragments, which is greatly facilitated by the use of the recently developed parallel pyrosequencing technology [11] (Figure 1). To concentrate the search on functionally important regions of the genome, the authors determined the positions of nucleosomes containing the histone variant H2A.Z, a histone typically enriched in nucleosomes at promoter regions.

Albert *et al.* [1] purified nucleosomes from the mixture of nucleosomes generated in a yeast strain in which the gene *htz1*, which encodes H2A.Z, had been epitope-tagged. Chromatin fragments containing H2A.Z nucleosomes were purified by chromatin immunoprecipitation and mono-nucleosomal DNA was isolated from the H2A.Z-enriched chromatin by gel purification. Altogether, 322,000 reads of DNA from nucleosomes enriched in the histone variant H2A.Z were made.

As H2A.Z-enriched nucleosomes represent a small proportion of all nucleosomes, it proved possible to obtain many reads for most of the highly enriched nucleosomes [1]. This would be expected to result in a series of sequence reads for each nucleosome with a normal distribution about a predominant location. Sequence reads of the most heavily H2A.Z-enriched nucleosomes do indeed fit well to a normal distribution. Further confidence in the assignment of these positions can be gained from the observation that the centers of these distributions match up well when the data obtained from both DNA strands are compared: the observed median error is 4 bp.

Another important point to bear in mind is that even in a homogeneous population of yeast cells, the positions of nucleosomes are likely to vary from cell to cell. Although a single type of nucleosome organization may dominate in some parts of the genome, in others, several mutually exclusive organizations with equally favorable nucleosome-DNA interactions may be possible. Albert *et al.* [1] used the enhanced resolution of their technique to address this variability in nucleosome position. Decomposing their raw data with narrower normal distributions, they obtained evidence for several translational settings (that is, positions along the DNA) of an individual H2A.Z nucleosome that typically occur with a 10-bp periodicity, which corresponds to a full turn of the DNA double helix. This is in accord with the concept of rotational phasing, whereby the same face of the DNA double helix makes contact with binding sites that are regularly spaced around the surface of histone octamers.
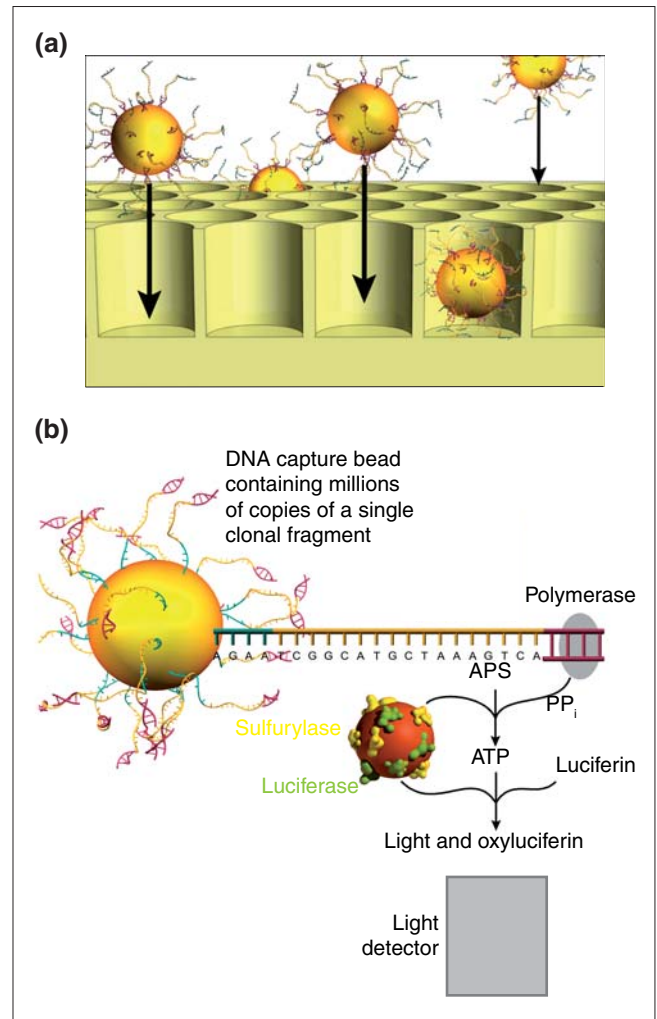
Figure 1
Automated ultrahigh-throughput parallel pyrosequencing. A single strand of each DNA fragment to be sequenced (a maximum of around 500 bp long) is amplified by PCR to make a single-strand DNA library and immobilized as one strand per bead. **(a)** Each bead is captured in its own microchamber and the DNA amplified by PCR, resulting in a bead coated with millions of copies of this single strand. **(b)** For sequencing, the DNA-coated beads are placed into a plate containing more than a million individual fiber-optic picoliter-scale wells, such that each well contains no more than one bead, and all the wells are then sequenced simultaneously by proprietary pyrosequencing technology. Pyrosequencing is based on primer-initiated DNA synthesis on the single strands, using repeated cycles of each of the four nucleotides (added as pyrophosphates) in turn. Addition of a base (or more than one base at a time) to the growing strand is accompanied by release of pyrophosphate and the consequent activation of luciferin and a flash of light, which is recorded via the fiber optics (the intensity of the flash correlates with the number of bases added). The light generation system depends on the reaction of pyrophosphate with APS [23] catalyzed by sulfurylate to give ATP. The luciferase catalyzed reaction of ATP and luciferin generates a flash of light. The sequence of the DNA in each well is then recovered from the recorded data. Figure reproduced with permission from [23].

Even the new mapping may have reached a limit here: the resolution of this approach depends not only on the accuracy

of the sequencing step but also on the sensitivity with which MNase detects the edges of nucleosomes. MNase has some sequence specificity itself, and to reduce the influence of artifacts attributable to this, an algorithm can be applied. A more substantial problem, however, is that the boundaries of nucleosomes are inherently dynamic. The outer turns of DNA are prone to transient dissociation via a process known as site exposure [12,13]. In addition, nucleosome-binding proteins may associate with nucleosomes to increase the amount of DNA protected to more than the 146 bp of the canonical nucleosome. These factors are likely to contribute to the relatively broad distribution of fragment lengths that result from genomic MNase digestion (see Supplementary Figure 2b of [1]), which in turn makes it difficult to assign the location of the central base pair of the nucleosomal DNA, especially with the relatively short sequence reads used in this study.

Preferential locations of H2A.Z nucleosomes

The availability of nucleosome-positioning information at this resolution provides an opportunity to search for DNA sequences that are preferentially incorporated into H2A.Z-containing nucleosomes. Albert *et al.* [1] observed that AA/TT and GC dinucleotides are distributed around the 8,000 best-positioned nucleosomes with a striking 10-bp periodicity. This distribution is similar to that previously observed both *in vitro* and *in vivo* [8,9]. A difference from previously observed patterns is that a deficiency of AT dinucleotides was detected 3-4 bp from the nucleosome border [1]. Further studies will be required to determine whether this has been picked up as a result of the increased resolution of this approach or whether it is a unique feature of H2A.Z-containing nucleosomes. It is likely, however, that the high resolution of this new approach will provide an opportunity to gain new insights into the rules that determine which DNA sequences favor nucleosome occupancy.

The limitations notwithstanding, the power of the current sequencing technology is breathtaking. The genome-wide map of H2A.Z-enriched nucleosomes generated by Albert *et al.* [1] is available in a highly accessible format via the authors' website [14]. It has been possible to assign trends in the location of H2A.Z-containing nucleosomes at a number of different classes of genes [1]. At the majority of genes transcribed by RNA polymerase II, H2A.Z-enriched nucleosomes were [1] detected immediately downstream of the start site for transcription, beyond the nucleosome-depleted promoter region, with lower levels in the coding region. A similar arrangement was observed at genes transcribed by RNA polymerase I, but differed subtly at TATA-containing genes transcribed by polymerase II and at genes transcribed by polymerase III. Some of these observations can be considered extensions to previous studies of H2A.Z nucleosome localization [15-19], but with superior spatial resolution. In future, it may be equally important to assess the duration of H2A.Z nucleosome occupancy. Recent observations suggest that a subset of nucleosomes undergo rapid cycles of assembly and removal and that these include nucleosomes enriched in H2A.Z [20,21]. Indeed, the dynamic nature of H2A.Z nucleosomes may underlie some of the discrepancies among previous studies.

While the sequencing technology will no doubt be pushed further, even at its current level it offers an exciting new way to study the distribution of DNA sequences across genomes. As it becomes cheaper and more widely available, there seems no reason why it should not offer an alternative to high-density microarrays. While Albert *et al.* [1] used parallel sequencing for nucleosome mapping, the same approach could be used to monitor mRNA levels, to sequence small RNAs [22] or to sequence the products of chromatin immunoprecipitations in higher eukaryotes as well as in yeast. The dawn of the pyrosequencing era could represent the beginning of the end for high-density DNA arrays.

References

1. Albert I, Mavrich TN, Tomsho LP, Qi J, Zanton SJ, Schuster SC, Pugh BF: **Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome.** *Nature* 2007, **446:**572-576.
2. Noll M, Kornberg RD: **Action of micrococcal nuclease on chromatin and the location of histone H1.** *J Mol Biol* 1977, **109:**393-404.
3. Thoma F, Bergman LW, Simpson RT: **Nuclease digestion of circular TRP1ARS1 chromatin reveals positioned nucleosomes separated by nuclease-sensitive regions.** *J Mol Biol* 1984, **177:**715-733.
4. Yenidunya A, Davey C, Clark D, Felsenfeld G, Allan J: **Nucleosome positioning on chicken and human globin gene promoters *in vitro*.** *J Mol Biol* 1994, **237:**401-414.
5. Yuan GC, Liu YJ, Dion MF, Slack MD, Wu LF, Altschuler SJ, Rando OJ: **Genome-scale identification of nucleosome positions in *S. cerevisiae*.** *Science* 2005, **309:**626-630.
6. Pokholok DK, Harbison CT, Levine S, Cole M, Hannett NM, Lee TI, Bell GW, Walker K, Rolfe PA, Herbolsheimer E, *et al.*: **Genome-wide map of nucleosome acetylation and methylation in yeast.** *Cell* 2005, **122:**517-527.
7. Simpson RT: ***In vivo* methods to analyze chromatin structure.** *Curr Opin Genet Dev* 1999, **9:**225-229.
8. Satchwell SC, Drew HR, Travers AA: **Sequence periodicities in chicken nucleosome core DNA.** *J Mol Biol* 1986, **191:**659-675.
9. Segal E, Fondufe-Mittendorf Y, Chen L, Thastrom A, Field Y, Moore IK, Wang JP, Widom J: **A genomic code for nucleosome positioning.** *Nature* 2006, **442:**772-778.
10. Whitehouse I, Tsukiyama T: **Antagonistic forces that position nucleosomes *in vivo*.** *Nat Struct Mol Biol* 2006, **13:**633-640.
11. Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, Church GM: **Accurate multiplex polony sequencing of an evolved bacterial genome.** *Science* 2005, **309:**1728-1732.
12. Polach KJ, Widom J: **Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation.** *J Mol Biol* 1995, **254:**130-149.
13. Li G, Levitus M, Bustamante C, Widom J: **Rapid spontaneous accessibility of nucleosomal DNA.** *Nat Struct Mol Biol* 2005, **12:**46-53.
14. **Nucleosome maps of the yeast genome** [http://atlas.bx.psu.edu/yeast-maps/yeast-index.html]
15. Meneghini MD, Wu M, Madhani HD: **Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin.** *Cell* 2003, **112:**725-736.
16. Raisner RM, Hartley PD, Meneghini MD, Bao MZ, Liu CL, Schreiber SL, Rando OJ, Madhani HD: **Histone variant H2A.Z marks the**

5' ends of both active and inactive genes in euchromatin. *Cell* 2005, **123:**233-248.

17. Zhang H, Roberts DN, Cairns BR: **Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss.** *Cell* 2005, **123:** 219-231.

18. Guillemette B, Bataille AR, Gevry N, Adam M, Blanchette M, Robert F, Gaudreau L: **Variant histone H2A.Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning.** *PLoS Biol* 2005, **3:**e384.

19. Li B, Pattenden SG, Lee D, Gutierrez J, Chen J, Seidel C, Gerton J, Workman JL: **Preferential occupancy of histone variant H2AZ at inactive promoters influences local histone modifications and chromatin remodeling.** *Proc Natl Acad Sci USA* 2005, **102:** 18385-18390.

20. Dion MF, Kaplan T, Kim M, Buratowski S, Friedman N, Rando OJ: **Dynamics of replication-independent histone turnover in budding yeast.** *Science* 2007, **315:**1405-1408.

21. Mito Y, Henikoff JG, Henikoff S: **Histone replacement marks the boundaries of *cis*-regulatory domains.** *Science* 2007, **315:**1408-1411.

22. Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, Bartel DP, Kingston RE: **Characterization of the piRNA complex from rat testes.** *Science* 2006, **313:**363-367.

23. **454 Life Sciences: Enabling Technology** [http://www.454.com/enabling-technology/sequencing.asp]