is, occupancy) *in vitro*. As an example of problems inherent in this comparison, consider two well-positioned nucleosomes *in vivo* for which the *in vitro* map has 5 reads within 20 bp of the first nucleosome's center and 10 reads within 20 bp of the second nucleosome's center. The analysis of Zhang *et al.*[2] assigns a higher positioning correspondence score to the second nucleosome because of its greater number of reads. Suppose, however, that the second nucleosome in the *in vitro* map is flanked also by many additional reads outside the 20-bp window, whereas the first nucleosome has no other reads nearby. In this case, the analysis of Zhang *et al.*[2] is inconsistent with that study's own definition of positioning because, according to that definition, the first nucleosome in the *in vitro* map is highly positioned and the second is not, yet the second receives the higher positioning correspondence score. Moreover, in contrast to previous work by Struhl and colleagues[14], the calculation in Zhang *et al.*[2] subtracts the amount that is attributable to random chance without rescaling the remainder and thus cannot yield a result in which 100% of the *in vivo* nucleosome positions were explained by the *in vitro* data.

To obtain a more direct comparison of nucleosome positions between the *in vitro* and *in vivo* maps, we used two measures (see below) of positioning to separately assign discrete nucleosome positions in both maps. From these discrete positions, we then calculated the fraction of the positioned nucleosomes *in vivo* that are explainable by the positions adopted by nucleosomes *in vitro*—that is, the fraction of the positioned nucleosomes *in vivo* that is attributable to intrinsic nucleosome sequence preferences. One measure is essentially that used by Zhang *et al.*[2]; we refer to it as 'localization'. It defines the positioning at every base pair *i* as the number of nucleosome reads that fall within a 40-bp region centered on *i*, divided by the number of reads within 160 bp of *i* and then smoothed with a Gaussian. This is identical to the measure used by Zhang *et al.*[2] except that we smooth the results and use a 40-bp window instead of one of 20 bp. This is done to better accommodate the sparseness of the data (summarized above: the map of Zhang *et al.*[2] has only ~5 reads per 20-bp window) and the limited accuracy with which nucleosome centers are known (the distribution of nucleosome lengths that result from micrococcal nuclease digestion is much greater than 40 bp wide[6,17,18]). We also corrected their calculation to allow for the full possible range of answers (as in ref. 14). Despite these improvements to the Zhang *et al.*[2] metric, we consider it problematic because we find that different results are obtained by slight variations of its parameters. We therefore also introduced a second measure, based on simple Gaussian smoothing of the raw nucleosome read data. As intended in Zhang *et al.*[2], both measures assign favorable scores to highly positioned nucleosomes, regardless of whether those nucleosomes are very abundant in the cell population, or very rare. The results using these two metrics and our original occupancy metric are included in **Figure 1a**,**b**.

In summary, although significant differences do exist between the *in vitro* and *in vivo* nucleosome maps, as we had previously noted, the existing literature and comparisons using both our data and those of Zhang *et al.*[2] all show that the genome explicitly encodes many aspects of the *in vivo* nucleosome organization through the nucleosomes' intrinsic DNA sequence preferences.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Noam Kaplan[1], Irene Moore[2], Yvonne Fondufe-Mittendorf[2], Andrea J Gossett[3,4], Desiree Tillo[5], Yair Field[1], Timothy R Hughes[5–7], Jason D Lieb[3], Jonathan Widom[2] & Eran Segal[1,8]

[1]Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel. [2]Department of Biochemistry, Molecular Biology, and Cell Biology, Northwestern University, Evanston, Illinois, USA. [3]Department of Biology, Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. [4]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. [5]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. [6]Terrence Donnelly Centre for Cellular & Biomolecular Research, Toronto, Ontario, Canada. [7]Banting and Best Department of Medical Research, Toronto, Ontario, Canada. [8]Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, Israel. e-mail: t.hughes@utoronto.ca (T.R.H.), jlieb@bio.unc.edu (J.D.L.), j-widom@northwestern.edu (J.W.) or eran.segal@weizmann.ac.il (E.S.)

1. Kaplan, N. *et al. Nature* **458**, 362–366 (2009).
2. Zhang, Y. *et al. Nat. Struct. Mol. Biol.* 16, 847–852 (2009).
3. Satchwell, S.C., Drew, H.R. & Travers, A.A. *J. Mol. Biol.* **191**, 659–675 (1986).
4. Segal, E. *et al. Nature* **442**, 772–778 (2006).
5. Johnson, S.M., Tan, F.J., McCullough, H.L., Riordan, D.P. & Fire, A.Z. *Genome Res.* **16**, 1505–1516 (2006).
6. Albert, I. *et al. Nature* **446**, 572–576 (2007).
7. Lowary, P.T. & Widom, J. *J. Mol. Biol.* **276**, 19–42 (1998).
8. Widlund, H.R. *et al. J. Mol. Biol.* **267**, 807–817 (1997).
9. Sekinger, E.A., Moqtaderi, Z. & Struhl, K. *Mol. Cell* **18**, 735–748 (2005).
10. Fraser, R.M., Keszenman-Pereyra, D., Simmen, M.W. & Allan, J. *J. Mol. Biol.* **390**, 292–305 (2009).
11. Visnapuu, M.L. & Greene, E.C. *Nat. Struct. Mol. Biol.* **16**, 1056–1062 (2009).
12. Iyer, V. & Struhl, K. *EMBO J.* **14**, 2570–2579 (1995).
13. Field, Y. *et al. Nat. Genet.* **41**, 438–445 (2009).
14. Peckham, H.E. *et al. Genome Res.* **17**, 1170–1177 (2007).
15. Segal, E. & Widom, J. *Nat. Rev. Genet.* **10**, 443–456 (2009).
16. Thastrom, A., Bingham, L.M. & Widom, J. *J. Mol. Biol.* **338**, 695–709 (2004).
17. Field, Y. *et al. PLoS Comput. Biol.* **4**, e1000216 (2008).
18. Valouev, A. *et al. Genome Res.* **18**, 1051–1063 (2008).

Evidence against a genomic code for nucleosome positioning

Zhang *et al.* reply:

It has been proposed that there is a "genomic code for nucleosome positioning"[1] in which the pattern of nucleosome positions *in vivo* is determined primarily by the genomic DNA sequence and can be predicted. As experimental support for such "DNA-encoded nucleosome organization," Kaplan *et al.*[2] generated genome-wide maps of nucleosomes assembled *in vitro* with purified histones and concluded that these are highly similar to maps of nucleosomes *in vivo*[2]. However, in similar experiments, we concluded that "intrinsic histone-DNA interactions are not the major determinant of nucleosome positions *in vivo*,"[3] thereby arguing against a nucleosome positioning code. The originally submitted correspondence of Kaplan *et al.*, to which our response was written, was entitled "a genomic code for nucleosome positioning," and it disputed our analyses and

major conclusion. We are pleased to see that the current correspondence[4] has now reduced the role of nucleosome sequence preferences from "encoding" to "influencing" *in vivo* nucleosome organization and leaves the issue of a code "for others to debate."

The *in vitro* mapping data in both studies is quite similar, and there is agreement that intrinsic histone-DNA interactions contribute to certain aspects of nucleosome positioning *in vivo*. The implication that we argue against any biological role of intrinsic histone-DNA interactions is incorrect and indeed inconsistent with our work over the past 25 years[5–8]. Nevertheless, we do disagree on the following: (i) the use of nucleosome occupancy measurements to assess nucleosome positioning; (ii) the impact of systematic errors in nucleosome occupancy measurements that overestimate the similarity between *in vivo* and *in vitro* samples; (iii) the ability of *in vitro* assembled nucleosomes to recapitulate the striking *in vivo* nucleosomal pattern; and (iv) the meaning of a nucleosome code. An independent analysis[9] of the two key papers[2,3] has supported our viewpoint.

The concept that histones have DNA sequence preferences for nucleosome formation was established 25 years ago. In pioneering experiments involving the sequencing of nucleosomal DNA generated by micrococcal nuclease (MNase), the same technique used today, Horace Drew and Andrew Travers showed that nucleosomal DNA *in vivo* has strong rotational positioning with 10–base pair (bp) helical periodicity that is due to preferences for dinucleotides that face inwards or outwards with respect to the histones and optimize DNA bending[10,11]. Around the same time, it was shown that poly(dA:dT) disfavors nucleosome formation *in vitro*[12,13] and increases chromatin accessibility *in vivo* via its intrinsic DNA structure, particularly at yeast promoter regions where these sequences are highly enriched[5–7]. Indeed, poly(dA:dT) and (to a lesser extent) dinucleotide frequencies are the most important factors in the algorithm of Kaplan *et al.*[2] for predicting nucleosome occupancy.

Prior to the initial paper proposing a nucleosome positioning code, a direct comparison of the location of nucleosomes assembled on the yeast *PET56-HIS3-DED1* region *in vivo* and *in vitro* with purified histones revealed that both promoter regions intrinsically disfavor nucleosome formation[8]. Furthermore, it was argued that DNA sequence is responsible for nucleosome depletion at most yeast promoter regions *in vivo*, based on genome-wide occupancy measurements *in vivo*. Specifically, the relative paucity of nucleosomes at promoter regions with respect to the corresponding coding regions is independent of transcriptional activity and hence is not due to activator– or RNA polymerase II elongation–dependent histone removal[8]. Thus, as the concept and specific aspects of how DNA sequence contributes to nucleosome location *in vivo* are well established and not at issue, the key disagreement is whether intrinsic histone-DNA interactions have the predominant role in setting up the *in vivo* pattern and thus constitute a code for nucleosome positioning.

Kaplan *et al.*[2] and Zhang *et al.*[3] extend the earlier comparison[8] of nucleosomes assembled *in vivo* and *in vitro* to the entire yeast genome, and at higher (in principle, nucleotide) resolution using high-throughput sequencing. In interpreting the resulting maps, a major conceptual issue concerns the difference between nucleosome 'occupancy' and 'positioning'. Nucleosome occupancy reflects the average histone levels on a given region of DNA in a population of cells, but it does not address where individual nucleosomes are positioned (that is, differently positioned nucleosomes within a genomic region all contribute to occupancy). In contrast, the translational position of an individual nucleosome refers to the specific 146-bp sequence covered by the histone octamer. On a population basis, positioning can range from perfect (all nucleosomes occupy a specific 146-bp stretch) to random (nucleosomes occupy all possible genomic positions equally). We did not criticize Kaplan *et al.*[2] for nucleosome occupancy measurements *per se* (indeed, we also made this useful measurement; see below for limitations) but rather for using occupancy measurements to infer nucleosome positioning. As acknowledged in their correspondence[4], Kaplan *et al.*[2] did not perform translational positioning analyses in their original paper. In addition, the independent validation experiment mentioned in the correspondence[4] uses a different method to measure histone occupancy[14], but it does not address translational positioning, the key point of disagreement.

Zhang *et al.*[3] explicitly examined translational positioning and the relationship between nucleosomes generated *in vivo* and *in vitro*, and we disagree with the correspondence[4] on this point. Specifically, we defined positioned nucleosomes from the *in vivo* mapping as 20-bp windows centered on the peak position on a gene-by-gene and location basis (+1, +2, etc. with respect to the mRNA initiation site). We then measured the percentage of nucleosome centers within these windows (100% being the value expected for perfect positioning) in the *in vitro* (and *in vivo*) data and compared this to randomly positioned nucleosomes (Fig. 4b–d of ref. 3). This analysis is unaffected by nucleosome centers flanking the 20-bp window, and hence the problematic example given in the correspondence[4] is incorrect and irrelevant (the issue raised does affect the genome-wide measurement of maximal nucleosome positioning degree in Fig. 4a of ref. 3, but this is not relevant to the direct comparison of *in vitro* and *in vivo* positions). We note that our analysis is restricted to nucleosomes that are well positioned *in vivo*, but the role of intrinsic histone-DNA interactions in setting up the striking *in vivo* pattern is the key biological issue. The analysis cannot be done on weakly positioned nucleosomes, as their locations are ill defined due to sequencing limitations.

Using *in vitro* data generated in either paper, we estimated that ~20% of the *in vivo* positioned nucleosomes are positioned due to intrinsic histone-DNA interactions. As done previously[15] and in contrast to the correspondence[4], this estimate involved an explicit correction for random chance occurrence. Our estimate is consistent both with the previous observation that 2 out of 7 *in vivo* positioned nucleosomes in the *PET56-HIS3-DED1* region were observed *in vitro*[8] and with a previous estimate of ~25% based on computational predictions of positioned nucleosomes[15].

In the correspondence[4], the authors performed a related positioning analysis using 40-bp windows and obtained a value of 34–41% (perhaps as high as 49% with unspecified data smoothing). However, the calculated values strongly depend on the input parameters and definitions, and the size of the window is particularly important. Indeed, we obtain a value of ~30% when using 40-bp windows (quite similar to that in the correspondence[4]) but only ~15% when using 10-bp windows. Conceptually, a positioned nucleosome has a unique location (1-bp window), and the operational reason for using larger windows is to account for incomplete or excessive trimming of nucleosomes by MNase, which is experimentally unavoidable. Hence, values at smaller window sizes are more meaningful for nucleosome positioning measurements, whereas larger window sizes (for example, 40 bp, or ~25% of all possible positions) begin to approach measurements of nucleosome occupancy (that is, all possible positions), not positioning. In addition, by reporting positioning measurements at each individual base pair (as opposed to restricting such

measurements to positioned nucleosomes), the correspondence[4] is essentially converting positioning information into nucleosome occupancy.

We agree with Kaplan *et al.*[2] that nucleosome occupancy is an important concept, and indeed the central conclusion of our earlier work[8] is that "intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast." However, we disagree that nucleosome occupancy can be measured simply by counting nucleosome reads. In particular, the use of MNase and Illumina sequencing introduces systematic errors in the measurements and results in overestimates of the similarity between *in vivo* and *in vitro* samples. Illumina sequencing shows systematic differences in DNA sequence coverage depending on base composition and causes artifactually high correlations between samples[16]. Indeed, although Kaplan *et al.*[2] emphasize a correlation of 0.74 between their *in vitro* and *in vivo* samples, Stein *et al.*[9] have shown that the correlation is only 0.3 when their *in vitro* sample is compared to an *in vivo* sample analyzed by high-resolution microarrays. MNase has well-known DNA sequence specificity[17], and this influences both the relative cleavage of linker regions and the relative cleavage of nucleosomal regions as a function of MNase concentration[3,18]. We agree with Kaplan *et al.*[2] that other parameters, notably sparseness of data, might lead to an underestimation of the correlation, but this issue has not been investigated.

Kaplan *et al.*[2] does not have an explicit control for either DNA sequencing or for sequence specificity of MNase cleavage, making it difficult to determine the extent to which these issues affect the correlation between their *in vivo* and *in vitro* samples. Zhang *et al.*[3] analyzed a sonicated control sample to assess DNA sequencing effects and observed a correlation of 0.15–0.2. In a recently performed control, we observed a correlation of 0.3 between MNase-digested naked DNA with all *in vitro* and *in vivo* nucleosomal samples, and this may be an underestimate due to sparseness of data. Thus, not only do nucleosome occupancy measurements not address nucleosome positioning, but methodological considerations also significantly reduce the correlation between *in vivo* and *in vitro* nucleosome occupancy.

Aside from the technical issues raised above, both studies agree that *in vitro* assembled nucleosomes do not show the striking *in vivo* pattern in which the +1 nucleosome centered just downstream from the mRNA initiation site is highly positioned, with more downstream nucleosomes arrayed in the coding region becoming gradually less positioned[19,20]. This pattern is the hallmark of 'statistical positioning' of nucleosomes from a fixed barrier such as a DNA-binding protein[21] or perhaps a nucleosome-free region[20]. Kaplan *et al.*[2] correctly argue that the low and nonphysiological level of histones in their *in vitro* assembly reaction is unsuitable for forming nucleosome arrays and hence observing statistical positioning, but this issue does not apply to Zhang *et al.*[3], where the histone:DNA ratio was physiological and nucleosome arrays clearly evident. The use of limiting histone concentrations by Kaplan *et al.*[2] is advantageous for measuring intrinsic affinities of different genomic regions. In this regard, differences in nucleosome positioning and occupancy between the two studies are of potential interest, although they do not affect the key issues discussed here.

The mechanism by which the +1 nucleosome is positioned is the key to understanding how the *in vivo* nucleosomal pattern is generated. *In vitro*, correctly localized +1 nucleosomes are formed only to a limited extent, and unlike the situation *in vivo*, the +1 nucleosome behaves similarly to all other nucleosomes (to +10) with respect to the degree of localization. In contrast, as shown by Zhang *et al.*[3], the position of the +1 nucleosome *in vivo* is strikingly linked to the location of the mRNA initiation site and preinitiation complex in both yeast and flies, arguing for a transcription-based mechanism. A transcription-based mechanism for positioning the +1 (and more downstream) nucleosomes is further supported by the observation that the barrier for the *in vivo* pattern of statistical positioning occurs specifically at promoters (as opposed to terminator regions that also appear to be depleted of nucleosomes) and is unidirectional (only in the downstream direction)[3]. Lastly, the loss of RNA polymerase II significantly alters nucleosome positioning to more closely match *in vitro* preferences, arguing for an important role of transcription in determining nucleosome positioning *in vivo*[18]. These observations are in striking contrast to the transcription-independent depletion of nucleosomes at yeast promoter regions with respect to their corresponding coding regions[8]. Further, these observations are inconsistent with the idea that intrinsic histone-DNA interactions are central to establishing where nucleosomes are actually positioned (as opposed to being absent) *in vivo*, and the correspondence[4] does not address these inconsistencies.

Lastly, we do not agree with the use of the terms "nucleosome code" and "DNA-encoded nucleosome organization" to describe the experimental observations, and indeed, these terms are not clearly defined[9]. In common parlance, a code involves a system of words, letters or symbols that convey definite meanings. The genetic code, by which nucleic acid sequence is translated into protein sequence with high accuracy, clearly fits this definition. In contrast, 15–40% similarity (depending on definitions and methods, and we believe that the lower values are more relevant for positioning) between *in vitro* and *in vivo* nucleosome positions clearly does not convey a definite meaning for DNA sequence. More generally, 'preferences' are conceptually different from 'codes'. Thus, although intrinsic histone-DNA interactions contribute, they are not the major determinant of nucleosome positions *in vivo*. As such, the proposed nucleosome code is not supported.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Yong Zhang[1], Zarmik Moqtaderi[2], Barbara P Rattner[3], Ghia Euskirchen[4], Michael Snyder[4], James T Kadonaga[3], X Shirley Liu[1] & Kevin Struhl[2]

[1]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard School of Public Health, Boston, Massachusetts, USA. [2]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts, USA. [3]Section of Molecular Biology, University of California, San Diego, La Jolla, California, USA. [4]Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, Connecticut, USA.
e-mail: kevin@hms.harvard.edu

1. Segal, E. *et al. Nature* **442**, 772–778 (2006).
2. Kaplan, N. *et al. Nature* **458**, 362–366 (2009).
3. Zhang, Y. *et al. Nat. Struct. Mol. Biol.* **16**, 847–852 (2009).
4. Kaplan, N. *et al. Nat. Struct. Mol. Biol.* **17**, 918–920 (2010).
5. Struhl, K. *Proc. Natl. Acad. Sci. USA* **82**, 8419–8423 (1985).
6. Chen, W., Tabor, S. & Struhl, K. *Cell* **50**, 1047–1055 (1987).
7. Iyer, V. & Struhl, K. *EMBO J.* **14**, 2570–2579 (1995).
8. Sekinger, E.A., Moqtaderi, Z. & Struhl, K. *Mol. Cell* **18**, 735–748 (2005).
9. Stein, A., Takasuka, T.E. & Collings, C.K. *Nucleic Acids Res.* **38**, 709–719 (2010).
10. Drew, H.R. & Travers, A.A. *J. Mol. Biol.* **186**, 773–790 (1985).
11. Satchwell, S.C., Drew, H.R. & Travers, A.A. *J. Mol. Biol.* **191**, 659–675 (1986).
12. Kunkel, G.R. & Martinson, H.G. *Nucleic Acids Res.* **9**, 6869–6888 (1981).
13. Prunell, A. *EMBO J.* **1**, 173–179 (1982).
14. Visnapuu, M.L. & Greene, E.C. *Nat. Struct. Mol. Biol.* **16**, 1056–1062 (2009).
15. Peckham, H.E. *et al. Genome Res.* **17**, 1170–1177 (2007).
16. Harismendy, O. *et al. Genome Biol.* **10**, R32 (2009).

17. Horz, W. & Altenburger, W. *Nucleic Acids Res.* **9**, 2643–2658 (1981).
18. Weiner, A., Hughes, A., Yassour, M., Rando, O.J. & Friedman, N. *Genome Res.* **20**, 90–100 (2010).
19. Yuan, G.-C. *et al. Science* **309**, 626–630 (2005).
20. Mavrich, T.N. *et al. Genome Res.* **18**, 1073–1083 (2008).
21. Fedor, M.J., Lue, N.F. & Kornberg, R.D. *J. Mol. Biol.* **204**, 109–127 (1988).

A preoccupied position on nucleosomes

To the Editor:

'Occupancy' is a measure of histone or nucleosome density. Occupancy is typically measured on a genomic scale using microarrays or through deep sequencing (**Fig. 1**). Kaplan *et al.*[1] were correct in that the underlying DNA sequence has a predominant influence on occupancy levels *in vivo*. However, this and related work[1–3] often interchanged the term 'occupancy' with 'positioning', which is confusing. 'Positioning' is a measure of the extent to which a population of nucleosomes resists deviating from its consensus location along the DNA and can be thought of in terms of a single reference point on the nucleosome, like its dyad (**Fig. 1**)[4]. A low standard deviation means high positioning. Zhang *et al.*[5,6] were correct in that the underlying DNA sequences are not widespread determinants of nucleosome positioning *in vivo*, although they are major determinants at some positions. An important question now is how nucleosomes become uniformly spaced and precisely positioned *in vivo*.

COMPETING FINANCIAL INTERESTS
The author declares no competing financial interests.

B Franklin Pugh

Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, Pennsylvania, USA.
e-mail: bfp2@psu.edu

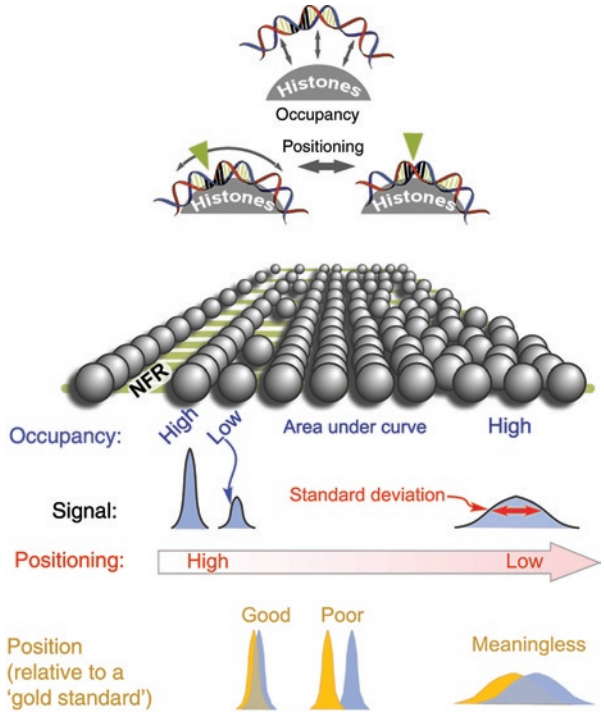

Figure 1 Illustration of how nucleosome occupancy and positioning differ. The upper panel shows a cross-section of a nucleosome, in which occupancy is distinguished from positioning. The lower panel shows how the two are measured. Occupancy is the area under the curve and reflects the local density of nucleosomes in a population, as illustrated by the column of spheres. Positioning or fuzziness is reflected in the standard deviation of the curve and is illustrated by how well the spheres are aligned in a column. The position of a nucleosome relative to some standard is indicated by how closely two peaks are separated. Comparing peaks of curves having high standard deviations is not likely to be meaningful because both peak locations have very high uncertainty.

1. Kaplan, N. *et al. Nature* **458**, 362–366 (2009).
2. Segal, E. *et al. Nature* **442**, 772–778 (2006).
3. Kaplan, N. *et al. Nat. Struct. Mol. Biol.* **17**, 918–923 (2010).
4. Albert, I. *et al. Nature* **446**, 572–576 (2007).
5. Zhang, Y. *et al. Nat. Struct. Mol. Biol.* **16**, 847–852 (2009).
6. Zhang, Y. *et al. Nat. Struct. Mol. Biol.* **17**, 920–923 (2010).