

# Correspondence: Reply to 'DNA shape is insufficient to explain binding'

Sivakanthan Kasinathan<sup>1,2,3</sup>, Gabriel E. Zentner<sup>4</sup>, Beibei Xin<sup>5</sup>, Remo Rohs<sup>5</sup> & Steven Henikoff<sup>1,6</sup>

*Nature Communications* 8:15644 doi: 10.1038/ncomms15644 (2017); Published 5 Jun 2017

Transcription factors (TFs) are DNA-binding proteins that regulate gene expression. Sequence-specific TFs recognize DNA via specific amino acid-base hydrogen bonds and contacts that read local DNA shape<sup>1</sup>. Studying base and shape readout modes of TFs *in vivo* has been challenging due to technical issues associated with current approaches for mapping TF-binding sites (TFBSs). We recently introduced Chromatin Endogenous Cleavage with sequencing (ChEC-seq), an *in vivo* mapping method based on fusing Micrococcal Nuclease (MNase) to a TF (ref. 2). Upon addition of calcium to permeabilized cells, tethered MNase cuts DNA adjacent to the bound TF and the released fragments are sequenced to provide a high-resolution genome-wide TFBS map. We used ChEC-seq to map the budding yeast TFs Abf1, Reb1 and Rap1 and obtained data similar to high-resolution ChIP-seq without the need for cross-linking, chromatin solubilization or antibodies.

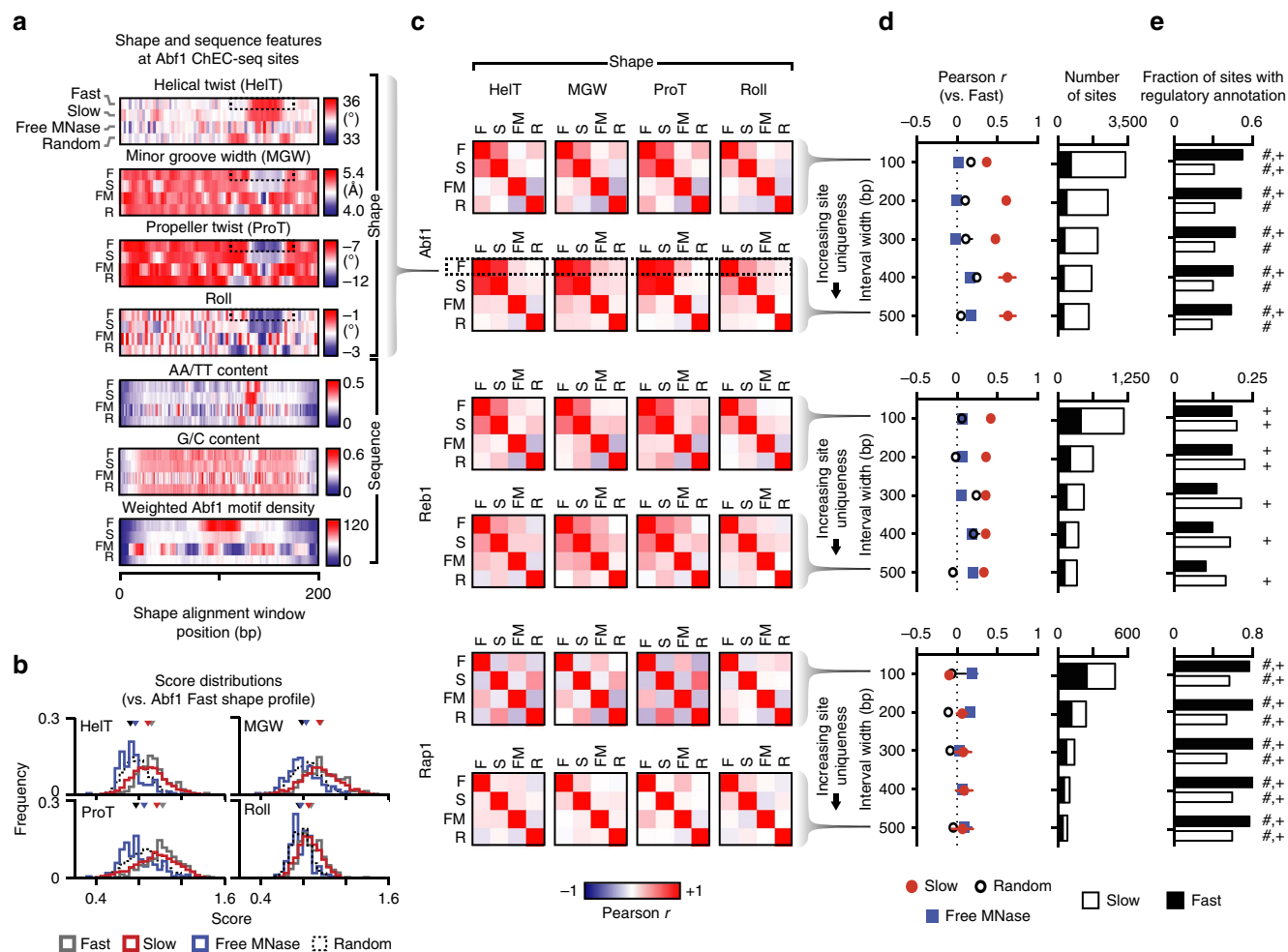
When cells were collected < 1 min after calcium addition, most TFBSs contained a TF-specific sequence motif ('fast' sites). We also reported 'slow' sites with low motif scores that appeared after ~10 min. We found that DNA shape features of high-scoring (mostly fast) and low-scoring (mostly slow) TFBSs corresponded closely, but differed from randomly chosen sites not overlapping high- or low-scoring sites. In our study, DNA shape features of fast and slow sites were centred on the best match to the TF consensus motif; however, randomly chosen genomic intervals were not similarly centred on the best motif match. Rossi, Lai and Pugh now find that when random sites are motif-centred, the shape features correspond closely to slow site features<sup>3</sup>, which might suggest that DNA shape is insufficient to explain binding site selection by the TFs Abf1, Reb1 and Rap1. However, given that sequence and shape features covary<sup>4</sup>, it is problematic to rely on motif-dependent analyses to draw conclusions about whether a TF recognizes DNA shape<sup>5</sup>.

To address this problem, we aligned DNA shape feature vectors for unique fast and slow ChEC-seq sites for each TF using a procedure that relied only on shape data and was not directly informed by sequence alignment. Given the possibility for overlap between nearby TFBSs, we identified unique sites that do not

intersect with any other ChEC-seq sites within intervals ranging from 100 to 500 bp surrounding ChEC-seq peak maxima, with larger windows associated with increasing stringency. For Abf1 and Reb1, we found that average fast and slow site shape features were well correlated at a range of interval widths ( $P \ll 0.001$ ; Fig. 1a–c). We also searched sites using a 'shape profile' defined using the average fast site features and found that score distributions for fast and slow sites only slightly differed ( $P > 0.03$ ), but were very different from random and free MNase sites ( $P \ll 10^{-10}$ ) for Abf1 (Fig. 1b) and Reb1 (not shown). The major shape feature proximal to Abf1 motifs is a deformation to the helix indicative of motif-proximal poly(dA:dT) tracts (Fig. 1a), a sequence feature we observed at slow sites in our original study<sup>2</sup>. Consistent with the recognition of a preferred shape signature by Abf1 and Reb1 at fast and slow sites, random sites and free MNase sites were not well correlated with fast and slow sites (Fig. 1a–c). We do not observe shape features enriched for poly(dA:dT) tracts at free MNase sites (Fig. 1a,b), suggesting that the detection of this shared shape feature at fast and slow ChEC-seq sites is not simply due to the higher prevalence of these features within nucleosome-depleted regions. Shape features at Rap1 fast and slow sites were not well correlated ( $P < 0.1$ ; Fig. 1c,d). The robustness of the correlation between average fast and slow shape features for Abf1 and Reb1 across a range of interval widths (Fig. 1d) suggests that sampling of similar shapes by TFs may explain binding events, even within promoters where fast and slow sites co-occur. From these motif-independent analyses, we conclude that fast and slow binding sites for Abf1 and Reb1 have similar shape features.

We next queried a TF-gene regulatory association database<sup>6</sup>, and asked whether TF-slow site associations had been previously observed in mapping or gene expression studies orthogonal to ChEC-seq. Consistent with our previous demonstration that slow sites were recovered as sites without the canonical motif in other studies<sup>2</sup>, the proportion of fast and slow sites documented or proposed to regulate proximal genes in previous studies (Fig. 1e) was similar across a range of interval widths. This suggests that slow sites with shape features similar to fast sites are likely true

<sup>1</sup>Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. <sup>2</sup>Medical Scientist Training Program, University of Washington School of Medicine, Seattle, Washington 98195, USA. <sup>3</sup>Molecular and Cellular Biology Graduate Program, University of Washington, Seattle, Washington 98195, USA. <sup>4</sup>Department of Biology, Indiana University, Bloomington, Indiana 47405, USA. <sup>5</sup>Molecular and Computational Biology Program, University of Southern California, Los Angeles, California 90089, USA. <sup>6</sup>Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. Correspondence and requests for materials should be addressed to S.H. (email: steveh@fredhutch.org).



**Figure 1 | Slow ChEC-seq sites have characteristic shapes separate from TF-binding motifs.** (a) Average shape and sequence features for unique fast (F) and slow (S) Abf1 sites 500 bp from other ChEC sites compared to free MNase (FM) and random (R) control sites aligned using shape features. Motif density was computed by weighting occurrences based on motif score. (b) Distributions of scores from searching shape vectors using the Abf1 fast site shape profile for fast and slow Abf1 sites (500 bp from other ChEC sites) and free MNase and random control sites; triangles represent the median score for each distribution. The shape profile used for searching is indicated in the dotted boxes in a. (c) Pearson correlations of average DNA helix twist (HelT), minor groove width (MGW), propeller twist (ProT), and roll features for unique fast and slow Abf1, Reb1 and Rap1 sites 100 bp (top) or 500 bp (bottom) from other ChEC sites compared to control sites aligned using shape features. (d) Pearson correlations compared to fast sites of aligned average DNA shape features at a variety of interval widths and the number of sites at each interval width. Error bars represent mean  $\pm$  s.e.m. (e) Proportion of unique sites at a range of overlap interval widths with known or proposed regulatory associations.  $P$  values  $< 0.1$  under Fisher's exact test versus FM and R control sites are indicated with '#' and '+', respectively.

binding sites and not simply experimental noise due to cleavage proximal to fast sites.

What accounts for the differential sensitivity of these TFs to DNA shape? All three TFs are essential and have roles in maintaining nucleosome organization<sup>7,8</sup>; however, Rap1 is unique in that it also functions in chromatin silencing at the mating type locus and telomeres<sup>9</sup>. Promoter architecture in *Saccharomyces cerevisiae* may provide a basis for this functional specialization<sup>4</sup>. We observed marked deviations in DNA shape in the average aligned fast and slow site profiles for Abf1 (Fig. 1a) and Reb1, but not Rap1 (not shown) consistent with the presence of poly(dA:dT) tracts, which are known to exclude nucleosomes and play a role in establishing canonical chromatin architecture<sup>4,10</sup>. Abf1 and Reb1 have been proposed to be dependent on poly(dA:dT) tracts for their localization and function<sup>11–13</sup>. It has been suggested that poly(dA:dT) tracts may participate in regulating ribosomal protein gene promoters, which are also bound by Rap1 (ref. 14); however, our inability to detect significant DNA shape contributions to Rap1 binding may

be due to the comparatively small number of sites tested. We speculate that promoters with poly(dA:dT) tracts not only exclude nucleosomes, but also have shape features that help recruit TFs that actively maintain nucleosome depletion<sup>15</sup>. Indeed, binding site-proximal poly(dA:dT) tracts have been proposed to enhance binding<sup>16</sup>, potentially by increasing accessibility of the adjacent major groove<sup>17</sup>. Thus, TF functional diversity and architecture of yeast promoters may explain the varying sensitivities of TFs to DNA shape. In this context, we anticipate that ChEC-seq will be a useful tool for generating high-resolution maps of protein-DNA interactions, with the potential to provide insights into the *in vivo* role of DNA shape in TFBS recognition.

## Methods

We defined unique sites such that the intersection of intervals of 100–500 bp widths centred on unique Abf1, Reb1, Rap1 and Free MNase ChEC-seq peak maxima was disjoint. As a null set, we generated 1,500 random intervals from the *sacCer3* genome assembly that did not overlap with ChEC-derived peaks. Shape features in 201-bp windows centred on peak maxima were determined as described<sup>4</sup> using the

DNAShapeR package<sup>18</sup>. At each interval width for a given TF, sites that did not have overlapping shape alignment windows were selected for alignment. Motif-independent alignment involved comparing each site against every other site within a given class and determining the shift that maximized the cosine similarity. Within a class, all sites were aligned to an internal centroid, defined as the site with the smallest sum of squared cosine similarities versus all other sites. Sites were then shifted relative to the centroid and class-specific average features were computed. Pearson's *r* was used to quantify the similarity of average shape features between classes (reported *P* values are two-tailed) without shifting the average features relative to each other. Given the strong A/T MNase cleavage preference (not shown) in the 5-bp window centred on peak maxima, we excluded these positions from the alignment. Further, because shape readout likely occurs near the TFBS, the largest shift considered was 25 bp and alignment was limited to the 90-bp interval centred at the peak maximum. Parameters used for all site classes including the random and free MNase sites were identical. Shape profiles for Abf1 and Reb1 were defined as the regions in the average fast shape features with the largest information gain relative to shuffled sequences. Score distributions were generated by scoring the aligned fast, slow, free MNase and random sites in the same 90-bp interval used for shape alignment using correlation distance to the shape profile; Mann–Whitney *U*-tests were performed for pairwise comparisons of the resulting distributions. To determine whether putative TFBSs regulate nearby genes, we assigned them to their closest ( $\leq 1$  kb) genes and queried YEASTRACT<sup>6</sup>. Source code for these analyses is publicly available ([https://github.com/sivakasinathan/shape\\_align](https://github.com/sivakasinathan/shape_align)).

## References

- Rohs, R. *et al.* Origins of specificity in protein–DNA recognition. *Annu. Rev. Biochem.* **79**, 233–269 (2010).
- Zentner, G. E., Kasinathan, S., Xin, B., Rohs, R. & Henikoff, S. ChEC-seq kinetics discriminates transcription factor binding sites by DNA sequence and shape *in vivo*. *Nat. Commun.* **6**, 8733 (2015).
- Rossi, M. J., Lai, W. K. M. & Pugh, B. F. DNA shape is insufficient to explain binding. *Nat. Commun.* **8**, 15643 (2017).
- Zhou, T. *et al.* DNAShape: a method for the high-throughput prediction of DNA structural features on a genomic scale. *Nucleic Acids Res.* **41**, W56–W62 (2013).
- Krietenstein, N. *et al.* Genomic nucleosome organization reconstituted with pure proteins. *Cell* **167**, 709–721. e712 (2016).
- Teixeira, M. C. *et al.* The YEASTRACT database: an upgraded information system for the analysis of gene and genomic transcription regulation in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **42**, D161–D166 (2014).
- Hartley, P. D. & Madhani, H. D. Mechanisms that specify promoter nucleosome location and identity. *Cell* **137**, 445–458 (2009).
- Ganapathi, M. *et al.* Extensive role of the general regulatory factors, Abf1 and Rap1, in determining genome-wide chromatin structure in budding yeast. *Nucleic Acids Res.* **39**, 2032–2044 (2011).
- Shore, D. RAP1: a protean regulator in yeast. *Trends Genet.* **10**, 408–412 (1994).
- Struhl, K. & Segal, E. Determinants of nucleosome positioning. *Nat. Struct. Mol. Biol.* **20**, 267–273 (2013).
- Wu, R. & Li, H. Positioned and G/C-capped poly(dA:dT) tracts associate with the centers of nucleosome-free regions in yeast promoters. *Genome Res.* **20**, 473–484 (2010).
- Moreira, J. M., Remacle, J. E., Kielland-Brandt, M. C. & Holmberg, S. Datin, a yeast poly(dA:dT)-binding protein, behaves as an activator of the wild-type ILV1 promoter and interacts synergistically with Reb1p. *Mol. Gen. Genet.* **258**, 95–103 (1998).
- Goncalves, P. M. *et al.* Transcription activation of yeast ribosomal protein genes requires additional elements apart from binding sites for Abf1p or Rap1p. *Nucleic Acids Res.* **23**, 1475–1480 (1995).
- Reja, R., Vinayachandran, V., Ghosh, S. & Pugh, B. F. Molecular mechanisms of ribosomal protein gene coregulation. *Genes Dev.* **29**, 1942–1954 (2015).
- Iyer, V. & Struhl, K. Poly(dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure. *EMBO J.* **14**, 2570–2579 (1995).
- Af ek, A., Sela, I., Musa-Lempel, N. & Lukatsky, D. B. Nonspecific transcription-factor–DNA binding influences nucleosome occupancy in yeast. *Biophys. J.* **101**, 2465–2475 (2011).
- Levo, M. *et al.* Unraveling determinants of transcription factor binding outside the core binding site. *Genome Res.* **25**, 1018–1029 (2015).
- Chiu, T.-P. *et al.* DNAShapeR: an R/Bioconductor package for DNA shape prediction and feature encoding. *Bioinformatics* **32**, 1211–1213 (2016).

## Author contributions

S.K., G.E.Z. and S.H. designed the study; S.K. wrote the software and performed the motif-independent alignments and statistical analyses; B.X. and R.R. performed DNA shape analyses; S.K. wrote the paper with input from all authors.

## Additional information

**Competing interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Kasinathan, S. *et al.* Correspondence: Reply to ‘DNA shape is insufficient to explain binding’. *Nat. Commun.* **8**, 15644 doi: 10.1038/ncomms15644 (2017).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

  The Author(s) 2017