

PROCEEDINGS

Open Access

# Genome-wide analysis of transcription factor binding sites and their characteristic DNA structures

Zhiming Dai<sup>1</sup>, Dongliang Guo<sup>1</sup>, Xianhua Dai<sup>1</sup>, Yuanyan Xiong<sup>2,3\*</sup>

From 10th International Symposium on Bioinformatics Research and Applications (ISBRA-14) Zhangjiajie, China. 28-30 June 2014

## Abstract

**Background:** Transcription factors (TF) regulate gene expression by binding DNA regulatory regions. Transcription factor binding sites (TFBSs) are conserved not only in primary DNA sequences but also in DNA structures. However, the global relationship between TFs and their preferred DNA structures remains to be elucidated.

**Results:** In this paper, we have developed a computational method to generate a genome-wide landscape of TFs and their characteristic binding DNA structures in *Saccharomyces cerevisiae*. We revealed DNA structural features for different TFs. The structural conservation shows positional preference in TFBSs. Structural levels of DNA sequences are correlated with TF-DNA binding affinities.

**Conclusions:** We provided the genome-wide correspondences of TFs to DNA structures. Our findings will have implications in understanding TF regulatory mechanisms.

## Background

Proper control of gene expression is critical for the complex function of a living cell. Although gene expression can be regulated at multiple levels, one of the most important regulatory mechanisms is at the transcriptional level. The transcriptional program is dependent on binding of transcription factors (TFs) to the cis-acting regulatory elements in promoter and enhancer regions of genes. Transcription factors also regulate gene expression by recruiting coactivators and RNA polymerase II (RNA Pol II) to target genes [1]. TFs and their binding sites are thus fundamental to the regulation of gene expression.

TFs bind DNA in a sequence-specific manner. Binding sites of one TF share conserved (i.e. similar) primary sequence patterns in different target promoters. The conserved sequence patterns have been widely used to computationally identify transcription factor binding sites (TFBSs) [2-5]. However, the traditional one-dimensional

view of DNA sequence is oversimplified. The three-dimensional structure of DNA, which reflects the physicochemical and conformational properties of DNA, is critical for the packaging of DNA in the cell [6]. The structure of DNA has been recognized to be important for protein-DNA recognition [7,8].

DNA bending plays a role in the regulation of prokaryotic transcription [9]. DNA structure can be used as discriminatory information to identify core-promoter regions [10,11]. Specific replication-related proteins show a preference to bind curved DNA sequences [12]. DNA curvature is also involved in the binding of recombination-related proteins to DNA [13]. DNA structure in the human genome is more evolutionary constrained than the primary nucleotide sequence alone [14]. Moreover, the DNA structure-conserved regions correlate with non-coding regulatory elements, better than sequence-conserved regions identified solely on the basis of primary sequence [14].

Although primary nucleotide sequences determine three-dimensional structures of DNA, different DNA sequences might have similar DNA structures, one TF

\* Correspondence: mody0911@gmail.com

<sup>2</sup>State Key laboratory for Biocontrol, Sun Yat-Sen University, Guangzhou 510275, China

Full list of author information is available at the end of the article

might bind DNA with different primary sequence patterns but with similar DNA structures. Recently, several computational approaches have used DNA structural properties to identify TFBSs with modest success [15-20]. There are many DNA structural properties that potentially influence TF-DNA binding. Different TFs might prefer different DNA structural properties. However, the full relationship between TFs and their corresponding DNA structural properties remains to be elucidated. In this study, we evaluated DNA structure in terms of various physicochemical and conformational properties. We have developed a computational approach to derive the first genome-wide landscape of TFs and their featured binding DNA structures in budding yeast *Saccharomyces cerevisiae*. We found that a considerable number of TFs have distinct DNA structural preferences. These structural features show positional preferences in TFBSs.

## Results

### A compendium of DNA structural properties

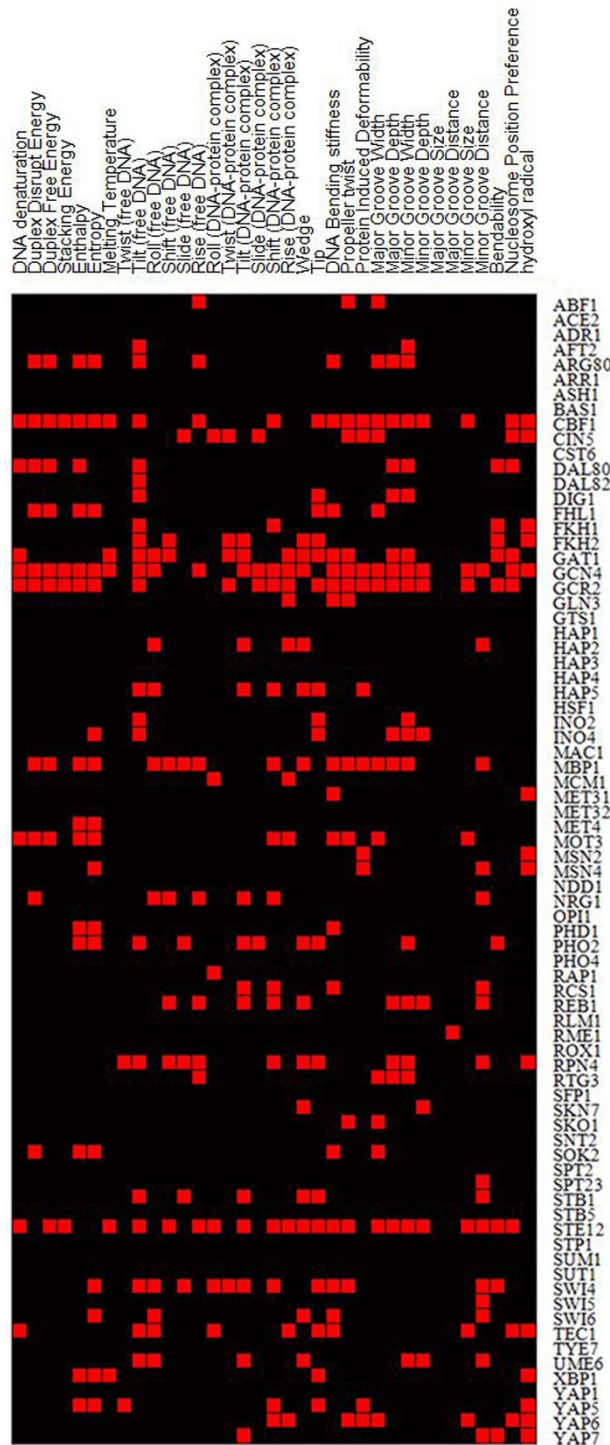
We used 35 types of di- or trinucleotide DNA structural properties, which were mainly collected in our previous study [21]. The structural properties chosen in this study have been frequently used and have been extensively studied in previous literatures [22,23]. These structural properties provide important information on the structure of DNA and capture structural properties that might be of importance for transcription. Each property contains complementary information and provides a unique insight into the DNA structure. The properties were classified into two types: conformational and thermodynamic. The rationale for exploiting di- or trinucleotide properties is the widely accepted nearest neighbor model saying that DNA structure can be understood and caused largely by interactions between neighboring base pairs [24,25]. This model is typically in the form of dinucleotide or trinucleotide properties. Each possible di- or trinucleotide and its reverse complement are assigned with a parametric value for a single structural property. The origins of the parametric values are either derived from experimentally determined structures, or from simulated structures of a DNA helix or a DNA-protein complex.

### Construction of the landscape of TFs and their characteristic binding DNA structures

We examined whether TFs show a preference to bind sequences with specific DNA structures. To this end, we examined whether binding sites of one particular TF are conserved in some DNA structures. We used genome-wide experimentally measured 6,390 TFBSs for 118 TFs in *S. cerevisiae* [26]. We restricted the analysis to TFs with more than 15 binding sites, resulting in 77 TFs. For each TF, we calculated the conservation rate in DNA structures of its TFBSs for each of the 35 DNA structural

properties (see Materials and Methods). DNA structure is dependent on DNA sequence. As TFBSs are known to be conserved in DNA primary sequences, this might bias the conservation of TFBSs in DNA structures. We should control conservation in DNA sequences when evaluating conservation in DNA structures. The conservation of TFBSs in DNA sequences could be measured by the information content (IC) of position weight matrices (PWMs) of TFBSs [27]. For each TF, we randomly generated a set of TFBSs from its real PWM, the number of which is the same as the number of its real TFBSs. The PWM of randomly generated TFBSs is the same as real PWM, so the conservation in DNA sequences of randomly generated TFBSs is the same as that of real TFBSs. We generated 10,000 randomized sets of TFBSs for each TF. For each set of TFBSs, we also calculated the conservation rate in DNA structure for each of the 35 DNA structural properties. For each TF, we calculated  $p$ -value for each structural property according to the ranking of its real conservation rate in those of 10,000 randomized sets. We found that 50 out of 77 (~65%) TFs bind DNA sequences that are significantly conserved in at least one structural property (ranging from one to twenty-six structural properties, a total of 356 pairs of TF-structure correspondences) ( $P < 0.05$ , after Bonferroni correction for multiple testing; Figure 1). This result indicates that a considerable number of TFs bind DNA sequences that show conservation in distinct DNA structures, independent of conservation in DNA sequences.

We next filtered the above landscape of TF-structure correspondences using more criteria. First, for each structural property, we randomly shuffled the parametric values among the di- or trinucleotides. We generated 10,000 randomized profiles for each structural property. For each TF, we calculated the conservation rates in DNA structures of its TFBSs as above based on these randomized profiles. For each TF, we calculated  $p$ -value for each structural property according to the ranking of its real conservation rate in those of 10,000 randomized profiles. If the 356 TF-structure pairs observed above is not an artifact, the real structural conservation rates of TFBSs should be significantly higher than those based on the randomized structural profiles. 39 out of the 356 TF-structure pairs show significantly higher conservation rates in the corresponding structures ( $P < 0.05$ , after Bonferroni correction for multiple testing). Second, the apparent conservation of TFBSs in DNA structures might be biased by the DNA structures of flanking regions around TFBSs. If TFBSs show similar DNA structural levels as their flanking regions, the conservation of TFBSs in DNA structures should be considered as an artifact. For the 39 pairs of TF-structure correspondences, we found 27 pairs whose TFBSs show significantly higher absolute levels in the corresponding



**Figure 1 The landscape of TFs and their characteristic binding DNA structures.** Rows represent TFs, and columns represent DNA structures. For each TF-structure pair, if structural conservation rate of its real TFBSs is significantly higher ( $P < 0.05$ , after Bonferroni correction for multiple testing) than those in 10,000 randomized experiments in which sequence conservation rates are the same as that of real TFBSs, it was colored red, otherwise it was colored black.

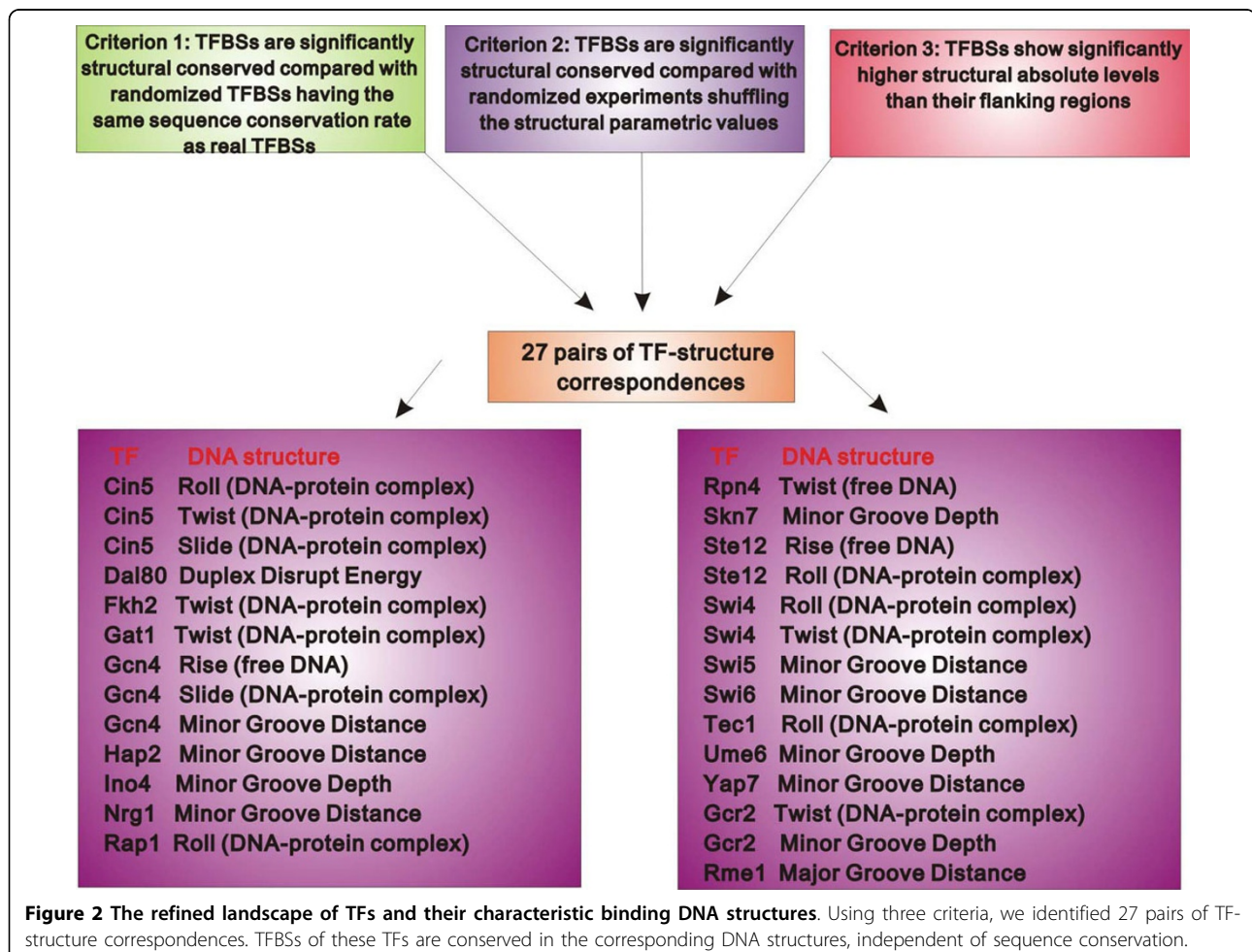
structures than their flanking regions (from -30 to +30 bp relative to TFBS) ( $P < 0.05$ , after Bonferroni correction for multiple testing). Together, we used three strict criteria to generate 27 pairs of TF-structure correspondences (Figure 2). We used these 27 TF-structure pairs in the following study unless otherwise stated.

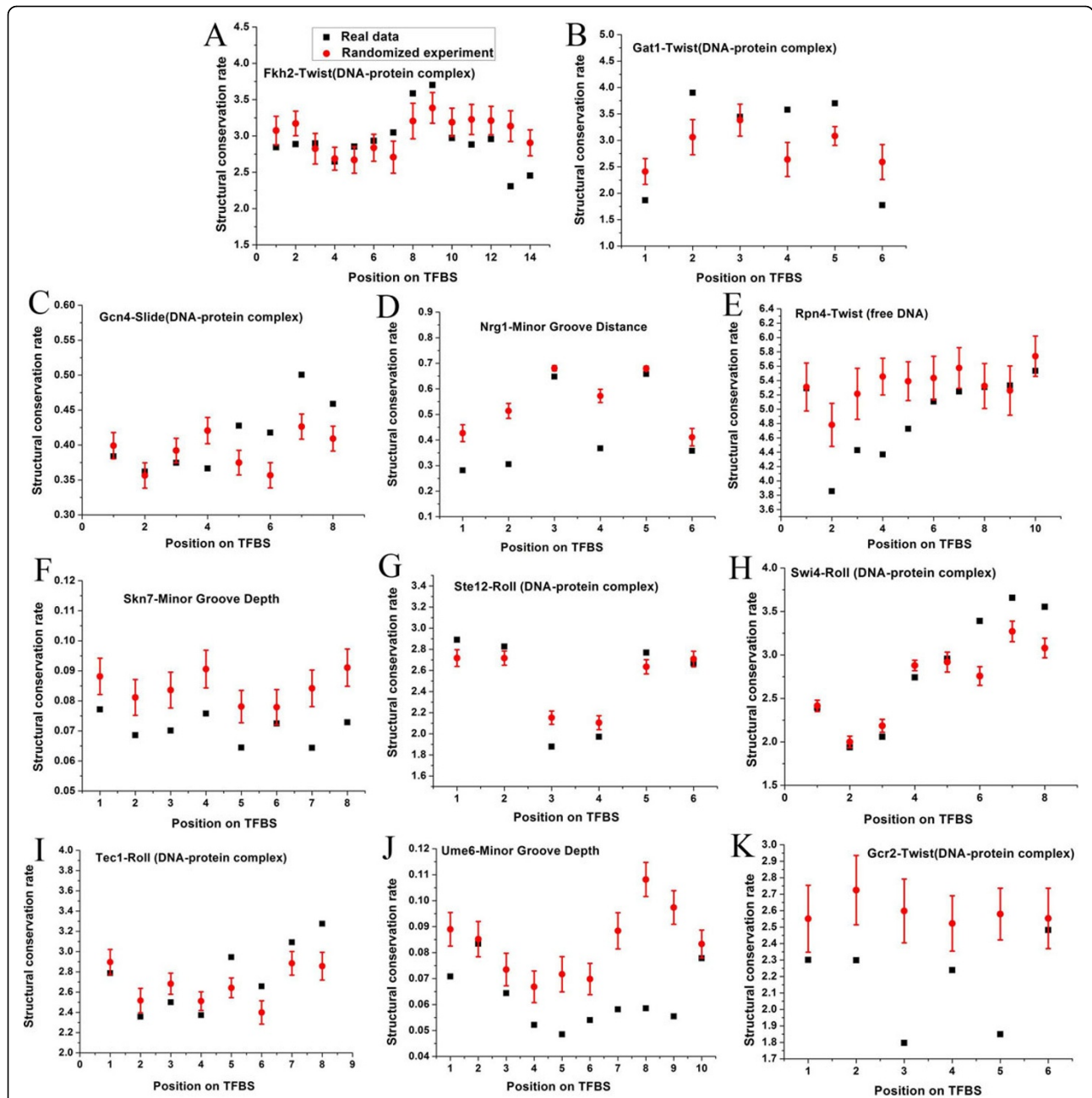
The 27 TF-structure pairs observed above demonstrate the characteristic associations between TFs and DNA structures of their binding sites. We found that there is selectivity of TFs and DNA structures involved in the associations: 20 of the 77 TFs examined show associations with DNA structures, and 9 of the 35 DNA structures examined are connected with TF binding (Figure 2). Furthermore, some specific TFs are associated with more DNA structures than the other TFs. There are two TFs (Cin5 and Gcn4) that are associated with three DNA structures.

### Structural conservation shows positional preferences in TFBSs

We asked whether TFs-associated structural conservation rates are homogeneous along TFBSs. To this end,

we compared DNA structural conservation rate of each position in TFBSs with those in 10,000 randomized experiments. As above, we used the random TFBSs generated from real PWMs. 11 out of 20 TFs listed in Figure 2 show significantly higher conservation in their correspondent structures in specific positions of TFBSs than those based on 10,000 randomized experiments ( $P < 0.05$ , after Bonferroni correction for multiple testing; Figure 3). The binding sites of most TFs show significantly higher structural conservation in more than one specific positions. The binding sites of two TFs, including Ste12 and Swi4, show significantly higher structural conservation in two successive positions. For example, conservation of roll property in the third and fourth positions of TF Ste12 binding sites (Figure 3G). These results suggest that DNA structures of some specific positions in TFBSs might be more important for the binding of TFs to DNA. For example, using an extensive categorization of the biophysical structures of TF DNA-binding domains [28,29], we found that Rap1 and Tec1, having the helix-turn-helix domains, show a preference





**Figure 3 Structural conservation shows positional preferences in TFBSs.** Real conservation rates of structures are shown for each position of TFBSs (black). Low levels correspond to high conservation rates. Average conservation rates of structures in 10,000 randomized experiments in which TFBSs are generated from real PWMs are also shown (red). Error bars were calculated by standard deviation. The names of TF-structure correspondences are also indicated. TFs show significantly higher conservation in their corresponding structures than those based on 10,000 randomized experiments in the following specific positions in TFBSs ( $P < 0.05$ , after Bonferroni correction for multiple testing): (A) The thirteenth and fourteenth positions; (B) The first and sixth positions; (C) The fourth position; (D) The first, second, third, fourth and fifth positions; (E) The second, third, fourth and fifth positions; (F) The first, second, third, fourth, fifth, seventh and eighth positions; (G) The third and fourth positions; (H) The third and fourth positions; (I) The third position; (J) The first, fourth, fifth, sixth, seventh, eighth and ninth positions; (K) The second, third, fourth and fifth positions.

to bind DNA sequences that are conserved in roll structural property.

#### TF-DNA binding affinities are correlated with DNA structural levels of binding sequences

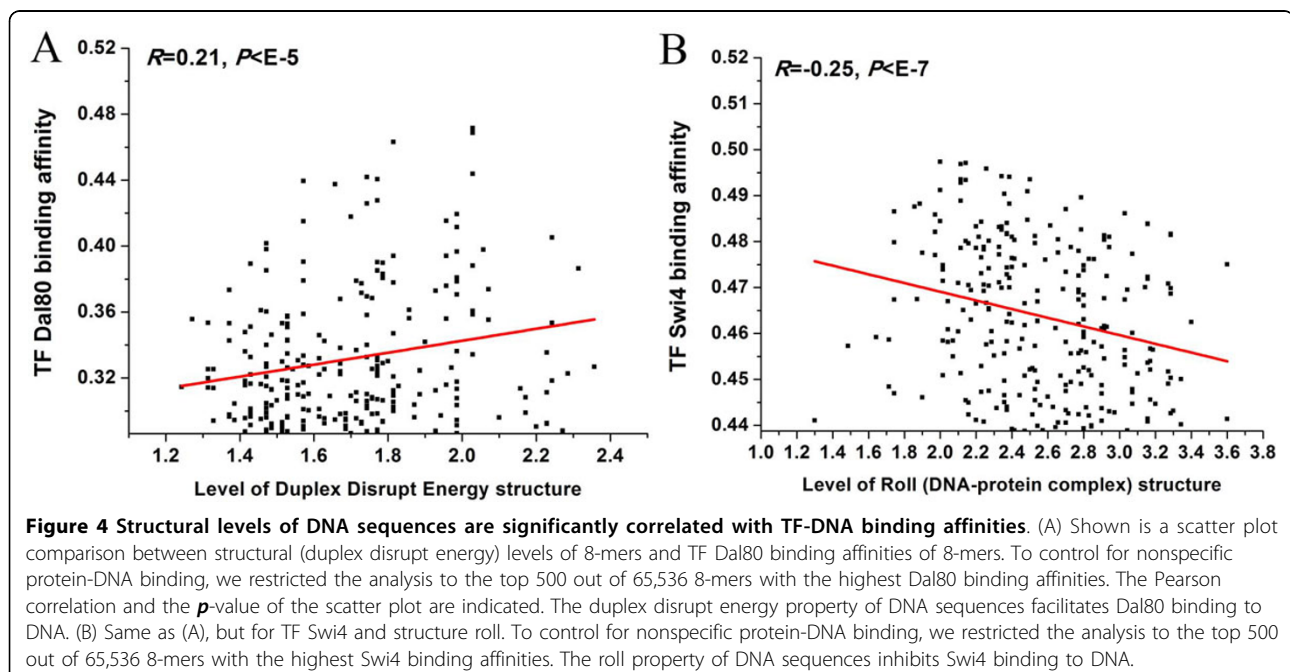
We asked whether TF-DNA binding affinities are correlated with DNA structures of binding sequences. A previous study has integrated binding affinities of 153 yeast TFs to all 8-bp sequences (8-mers) ( $N = 65,536$ ) *in vitro* utilizing protein-binding microarray (PBM) [30]. We used this data instead of *in vivo* data because *in vivo* TF-DNA binding is influenced by many factors besides TFBS, including nucleosome positioning, histone modification and so on. PBM data [30] is available for 14 out of 20 TF listed in Figure 2. For each 8-mer, we calculated its structural level for each of the 35 structural properties. We found that binding affinities of 10 out of 14 TFs to DNA are significantly correlated with their correspondent structural levels of DNA sequences (Pearson correlation coefficient,  $|R| > 0.1$ ,  $P < 0.05$ ; see selected examples in Figure 4). These results suggest that our observed TF-associated structures play a role in TF binding.

#### Discussion

In this study, we performed a systematic analysis to reveal the relationship between TFs and their preferred DNA structures. Using three strict criteria, we found that a considerable number of TFs bind DNA sequences that are structurally conserved, independent of sequence conservation in *S. cerevisiae*. Moreover, we found that

the structural conservation of TFBSs is also prevalent in other eukaryotes (unpublished data). These three strict criteria are very important to ensure a low level of false positives. However, some TFs do not show association with DNA structure. It does not indicate that DNA structure is not important to binding of these TFs to DNA. First, structural conservation of TFBSs might be largely determined by sequence conservation, so that structural conservation could not be detected when controlling for sequence conservation. Second, TFBSs of these TFs might be conserved in some unknown DNA structures. Advances in structural biology will give more insights into structures of TFBSs.

A key finding of this study is that structural conservation shows positional preference in TFBS. As our analysis is controlled for sequence conservation, the positional preference of structural conservation is not an artifact of the positional preference of sequence conservation. This finding could tell which position in TFBS is more important to TF-DNA binding. The local structure determined by these positions is more critical for TF-DNA recognition. The change in these local structure is more likely to influence TF-DNA binding and subsequent TF regulation. More attention should be paid to these local structures when analyzing cancer cell lines. It also will have implication in synthetic biology. It might help to distinguish functional TFBSs from non-functional TFBSs. On the other hand, some TFs whose binding sites are structurally conserved do not show structural positional preference. The binding of these TFs to DNA might be dependent on the DNA structure of the whole TFBS.



Despite its success, our approach has limitations. TFs generally interact with different protein factors to regulate target genes. These protein factors might influence the conformation of TFs, changing TF binding preference. TFs with similar DNA-binding domains might show different structural preferences for binding of DNA. One TF might even show different structural preferences for different target genes due to its different protein partners. Our method might miss this type of TF-structure correspondence.

## Materials and methods

### Calculation of DNA structural conservation rate

We used 35 types of conformational and thermodynamic DNA di- or trinucleotide structural properties, which were used in our previous study [21] (see Additional file 1 for more details about each of these structural properties), as measures of DNA structure. For a DNA region, the sequence is divided into overlapping di- or trinucleotide sequences. Structural profiles from DNA sequences are calculated for each structural property (except for hydroxyl radical cleavage pattern) as follows: The corresponding parametric value for each di- or trinucleotide was assigned to the first nucleotide of the di- or trinucleotide. In this way, the nucleotide sequence is converted into a sequence of numbers (i.e., a numerical profile). For hydroxyl radical cleavage intensity data, structural profiles are calculated as the reference where the data was published [31]. The hydroxyl radical cleavage intensity data are assigned to each nucleotide in each trinucleotide sequence. Note that the three nucleotides in each trinucleotide sequence have different values of hydroxyl radical cleavage intensity. As each nucleotide (except for the two terminal nucleotides at each end of the DNA region) is covered by three overlapping trinucleotide sequences, it has three values of hydroxyl radical cleavage intensity (one for each trinucleotide). The three values are averaged to produce hydroxyl radical cleavage intensity for each nucleotide. In this way, the nucleotide sequence is converted into a sequence of numbers (i.e., a numerical profile). For each region, the average of its numerical profile is considered as the level of the corresponding structure. For each pair of regions (e.g. TFBSs), we calculated the absolute difference values of structural profiles. For each TF, we calculated absolute difference profiles of structural profiles between every possible pairs of TFBSs (Additional file 2). We considered the average of resulting absolute difference profiles normalized by the length of TFBSs as a measure of conservation rate of DNA structure. The low values correspond to high conservation rates. In this way, there were 35 measures of structural conservation rate for TFBSs of each TF. Similarly, we also calculated absolute difference value of structural profiles at each position between every possible

pairs of TFBSs, and then calculated conservation rate of DNA structure at each position of TFBS.

### Data preparation

Transcription factor binding data was taken from MacIsaac et al. [26]. A *p*-value cutoff of 0.005 and conservation among three species was used to define the sequence bound by a particular TF. By applying this strict binding threshold, we ensured a low level of false positives. The data set includes 6,390 binding sites for 118 TFs. We mapped binding sites to the corresponding genes according to their located promoters (600 bp upstream of the gene in this study, the upstream region was truncated if it overlapped with neighboring genes). If the binding sites locate between divergent gene pairs, we mapped the binding sites to their nearest genes.

Gene coordinate data and genome sequence were downloaded from the *Saccharomyces* Genome Database [32]. TF binding affinity data for 8-mers were taken from Gordân et al. [30]. TF classification data were downloaded from two literatures [28,29].

### Statistical method

Given two samples of values, the Mann-Whitney U-test is designed to examine whether they have equal medians. The main advantage of this test is that it makes no assumption that the samples are from normal distributions.

## Additional material

**Additional file 1: Table S1** List of dinucleotide/trinucleotide DNA structural properties and their corresponding parameters

**Additional file 2: Figure S1** An example of how to calculate absolute difference profiles of structural profiles between one pair of TFBSs. For each TF, we calculated absolute difference profiles of structural profiles between every possible pairs of TFBS. We considered the average of resulting absolute difference profiles normalized by the length of TFBSs as a measure of conservation rate of DNA structure.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

ZD and DG implemented the algorithms and carried out the experiments. ZD also designed the study, analyzed the results and drafted the manuscript. DG, XD and YX participated in the analysis and discussion. All authors read and approved the final manuscript.

### Acknowledgements

We thank Qian Xiang and Shuaibin Lian for helpful discussions on the manuscript. The research has been supported by National Natural Science Foundation of China (NSFC) (Grant 61202343), by Natural Science Foundation of Guangdong Province (S2012040007935), and also by Fundamental Research Funds for the Central Universities (Grant 13lgpy06).

### Declarations

The research and publication has been supported by National Natural Science Foundation of China (NSFC) (Grant 61202343), by Natural Science

Foundation of Guangdong Province (S2012040007935), and also by Fundamental Research Funds for the Central Universities (Grant 13Jgpy06). This article has been published as part of *BMC Genomics* Volume 16 Supplement 3, 2015: Selected articles from the 10th International Symposium on Bioinformatics Research and Applications (ISBRA-14): Genomics. The full contents of the supplement are available online at <http://www.biomedcentral.com/bmcgenomics/supplements/16/S3>.

#### Authors' details

<sup>1</sup>Department of Electronics and Communication Engineering, School of Information Science and Technology, Sun Yat-Sen University, Guangzhou 510006, China. <sup>2</sup>State Key laboratory for Biocontrol, Sun Yat-Sen University, Guangzhou 510275, China. <sup>3</sup>SYSU-CMU Shunde International Joint Research Institute, Shunde, China.

Published: 29 January 2015

#### References

- Lelli KM, Slattery M, Mann RS: **Disentangling the many layers of eukaryotic transcriptional regulation.** *Annual review of genetics* 2012, **46**:43-68.
- Lawrence CE, Altschul SF, Boguski MS, Liu JS, Neuwald AF, Wootton JC: **Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment.** *Science* 1993, **262**(5131):208-214.
- Hertz GZ, Stormo GD: **Identifying DNA and protein patterns with statistically significant alignments of multiple sequences.** *Bioinformatics (Oxford, England)* 1999, **15**(7-8):563-577.
- Price A, Ramabhadran S, Pevzner PA: **Finding subtle motifs by branching from sample strings.** *Bioinformatics (Oxford, England)* 2003, **19** Suppl 2: ii149-155.
- Tompa M, Li N, Bailey TL, Church GM, De Moor B, Eskin E, Favorov AV, Frith MC, Fu Y, Kent WJ, et al: **Assessing computational tools for the discovery of transcription factor binding sites.** *Nature biotechnology* 2005, **23**(1):137-144.
- Olson WK, Gorin AA, Lu XJ, Hock LM, Zhurkin VB: **DNA sequence-dependent deformability deduced from protein-DNA crystal complexes.** *Proceedings of the National Academy of Sciences of the United States of America* 1998, **95**(19):11163-11168.
- Rohs R, West SM, Sosinsky A, Liu P, Mann RS, Honig BCINNO, Pmid: **The role of DNA shape in protein-DNA recognition.** *Nature* 2009, **461**(7268):1248-1253.
- Rohs R, West SM, Liu P, Honig B: **Nuance in the double-helix and its role in protein-DNA recognition.** *Current opinion in structural biology* 2009, **19**(2):171-177.
- Perez-Martin J, Rojo F, de Lorenzo V: **Promoters responsive to DNA bending: a common theme in prokaryotic gene expression.** *Microbiological reviews* 1994, **58**(2):268-290.
- Abeel T, Saeys Y, Bonnet E, Rouze P, Van de Peer Y: **Generic eukaryotic core promoter prediction using structural features of DNA.** *Genome research* 2008, **18**(2):310-323.
- Florquin K, Saeys Y, Degroevae S, Rouze P, Van de Peer Y: **Large-scale structural analysis of the core promoter in mammalian and plant genomes.** *Nucleic acids research* 2005, **33**(13):4255-4264.
- Ueguchi C, Kakeda M, Yamada H, Mizuno T: **An analogue of the DnaJ molecular chaperone in Escherichia coli.** *Proc Natl Acad Sci USA* 1994, **91**(3):1054-1058.
- Mazin A, Milot E, Devoret R, Chartrand P: **KIN17, a mouse nuclear protein, binds to bent DNA fragments that are found at illegitimate recombination junctions in mammalian cells.** *Molecular & general genetics: MGG* 1994, **244**(4):435-438.
- Parker SC, Hansen L, Abaan HO, Tullius TD, Margulies EH: **Local DNA topography correlates with functional noncoding regions of the human genome.** *Science (New York, NY)* 2009, **324**(5925):389-392.
- Broos S, Soete A, Hooghe B, Moran R, van Roy F, De Bleser P: **PhysBinder: improving the prediction of transcription factor binding sites by flexible inclusion of biophysical properties.** *Nucleic Acids Res* 2013, **41**:W531-534.
- Hooghe B, Broos S, van Roy F, De Bleser P: **A flexible integrative approach based on random forest improves prediction of transcription factor binding sites.** *Nucleic acids research* 2012, **40**(14):e106.
- Meysman P, Dang TH, Laukens K, De Smet R, Wu Y, Marchal K, Engelen K: **Use of structural DNA properties for the prediction of transcription-factor binding sites in Escherichia coli.** *Nucleic acids research* 2011, **39**(2): e6.
- Bauer AL, Hlavacek WS, Unkefer PJ, Mu F: **Using sequence-specific chemical and structural properties of DNA to predict transcription factor binding sites.** *PLoS computational biology* 2010, **6**(11):e1001007.
- Greenbaum JA, Parker SC, Tullius TD: **Detection of DNA structural motifs in functional genomic elements.** *Genome research* 2007, **17**(6):940-946.
- Maienschein-Cline M, Dinner AR, Hlavacek WS, Mu F: **Improved predictions of transcription factor binding sites using physicochemical features of DNA.** *Nucleic acids research* 2012, **40**(22):e175.
- Dai Z, Dai X: **Gene expression divergence is coupled to evolution of DNA structure in coding regions.** *PLoS Comput Biol* 2011, **7**(11):e1002275.
- Pedersen AG, Jensen LJ, Brunak S, Staerfeldt HH, Ussery DW: **A DNA structural atlas for Escherichia coli.** *Journal of molecular biology* 2000, **299**(4):907-930.
- Liao GC, Rehm EJ, Rubin GM: **Insertion site preferences of the P transposable element in Drosophila melanogaster.** *Proceedings of the National Academy of Sciences of the United States of America* 2000, **97**(7):3347-3351.
- Baldi P, Baisnee PF: **Sequence analysis by additive scales: DNA structure for sequences and repeats of all lengths.** *Bioinformatics (Oxford, England)* 2000, **16**(10):865-889.
- Goodsell DS, Dickerson RE: **Bending and curvature calculations in B-DNA.** *Nucleic Acids Res* 1994, **22**(24):5497-5503.
- Maclsaac KD, Wang T, Gordon DB, Gifford DK, Stormo GD, Fraenkel E: **An improved map of conserved regulatory sites for Saccharomyces cerevisiae.** *BMC bioinformatics* 2006, **7**:113.
- GuhaThakurta D: **Computational identification of transcriptional regulatory elements in DNA sequence.** *Nucleic acids research* 2006, **34**(12):3585-3598.
- Wingender E, Chen X, Hehl R, Karas H, Liebich I, Matys V, Meinhardt T, Pruss M, Reuter I, Schacherer F: **TRANSFAC: an integrated system for gene expression regulation.** *Nucleic acids research* 2000, **28**(1):316-319.
- Badis G, Chan ET, van Bakel H, Pena-Castillo L, Tillo D, Tsui K, Carlson CD, Gossett AJ, Hasinoff MJ, Warren CL, et al: **A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters.** *Molecular cell* 2008, **32**(6):878-887.
- Gordan R, Murphy KF, McCord RP, Zhu C, Vedenko A, Bulyk ML: **Curated collection of yeast transcription factor DNA binding specificity data reveals novel structural and gene regulatory insights.** *Genome biology* 2011, **12**(12):R125.
- Greenbaum JA, Pang B, Tullius TD: **Construction of a genome-scale structural map at single-nucleotide resolution.** *Genome research* 2007, **17**(6):947-953.
- Hirschman JE, Balakrishnan R, Christie KR, Costanzo MC, Dwight SS, Engel SR, Fisk DG, Hong EL, Livstone MS, Nash R, et al: **Genome Snapshot: a new resource at the Saccharomyces Genome Database (SGD) presenting an overview of the Saccharomyces cerevisiae genome.** *Nucleic Acids Res* 2006, **34**:D442-445.

doi:10.1186/1471-2164-16-S3-S8

**Cite this article as:** Dai et al.: Genome-wide analysis of transcription factor binding sites and their characteristic DNA structures. *BMC Genomics* 2015 **16**(Suppl 3):S8.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

