

ARTICLE SUMMARY

DNA-dependent formation of transcription factor pairs alters their binding specificity

Arttu Jolma, Yimeng Yin, Kazuhiro R. Nitta, Kashyap Dave, Alexander Popov, Minna Taipale, Martin Enge, Teemu Kivioja, Ekaterina Morgunova & Jussi Taipale*

PAPER ABSTRACT

Gene expression is regulated by transcription factors (TFs), proteins that recognize short DNA sequence motifs. Such sequences are very common in the human genome, and an important determinant of the specificity of gene expression is the cooperative binding of multiple TFs to closely located motifs. However, interactions between DNA-bound TFs have not been systematically characterized. To identify TF pairs that bind cooperatively to DNA, and to characterize their spacing and orientation preferences, we have performed consecutive affinity-purification systematic evolution of ligands by exponential enrichment (CAP-SELEX) analysis of 9,400 TF–TF–DNA interactions. This analysis revealed 315 TF–TF interactions recognizing 618 heterodimeric motifs, most of which have not been previously described. The observed cooperativity occurred promiscuously between TFs from diverse structural families. Structural analysis of the TF pairs, including a novel crystal structure of MEIS1 and DLX3 bound to their identified recognition site, revealed that the interactions between the TFs were predominantly mediated by DNA. Most TF pair sites identified involved a large overlap between individual TF recognition motifs, and resulted in recognition of composite sites that were markedly different from the individual TF's motifs. Together, our results indicate that the DNA molecule commonly plays an active role in cooperative interactions that define the gene regulatory lexicon.

SUMMARY

Charting transcription factor interactions. By using a new method, we discovered a large number of transcription factor (TF) pairs that bind cooperatively to DNA, and found that in many cases the TF pairs recognize a composite motif ('compound DNA word') that is markedly different from that expected from the individual TF motifs ('DNA words').

The problem

Individual TF recognition sequences (DNA words) are common in the human genome, so the specificity of gene expression depends on the cooperative binding of multiple TFs to sites close to each other. The critical role of particular TF combinations in cell-fate determination and development is well established^{1,2}. The TFs in cultured cells bind to only a subset of their potential target sites, and many of the occupied sites do not contain high-affinity TF motifs, suggesting that cooperative interactions allow TFs to bind to low-affinity sites^{3–5}. Research has shown many examples where two TFs bind DNA together as cooperative complexes that form as a result of protein–protein interactions⁶, DNA-facilitated protein–protein contacts, and interactions mediated by DNA^{7–9}. Most cases have been discovered by studying individual TF pairs, however, so it has not been known how common such TF–TF interactions are. The lack of understanding of TF interactions is one reason why the gene regulatory code, which determines how DNA sequence defines gene-expression patterns, has remained poorly understood.

The solution

To systematically identify TF–TF interactions in the presence of DNA, we developed a novel assay, consecutive affinity-purification systematic evolution of ligands by exponential enrichment (CAP-SELEX), in which two TFs are mixed with a diverse library of DNA fragments, after which sequences bound by both TFs are separated from the unbound or individually bound DNA molecules (see Graphical abstract). The selected DNA fragments are amplified by the polymerase chain reaction (PCR), and this new library of DNA fragments is used in another cycle of the same process. After sequencing the libraries, we identified cooperative TF–TF pairs by either detecting enrichment of a composite motif or by finding preferential arrangements (orientation and spacing preferences) between the two individual target motifs. A total of 9,400 pairwise combinations of TFs were tested using a set of 100 TFs against another set of 94 TFs. Our experiments indicated that 315 pairs of TFs bound DNA cooperatively. Structural modelling revealed that most of the cooperative interactions did not involve strong protein–protein contacts, but were facilitated or mediated by DNA. Surprisingly, the grammar of the genomic language was more complex than that of human languages. Instead of joining two words by deleting a space, the individual DNA

DETAILS

Citation: *Nature* **527**, 384–388 (2015)

doi: 10.1038/nature15518

● RECEIVED 23 January 2015

● ACCEPTED 24 August 2015

● PUBLISHED ONLINE 9 November 2015



*Department of Biosciences and Nutrition, Karolinska Institutet, SE 141 83, Sweden. Genome-Scale Biology Program, University of Helsinki, P.O. Box 63, FI-00014, Finland.

words were commonly altered when they formed compound DNA words, leading to a large expansion of the gene regulatory lexicon.

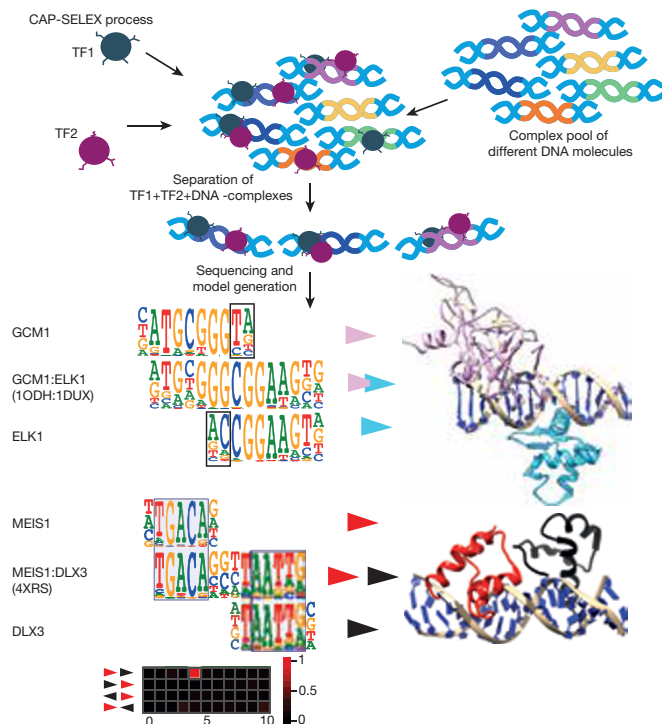
The implications

Our results show that cooperativity is an inherent feature of TF–DNA binding, and that DNA functions as an active interacting partner, commonly facilitating interactions between a wide range of TFs from diverse structural families. This DNA sequence-dependent mode of formation of TF pairs allows more interaction pairs than direct protein–protein interactions, as it prevents one highly expressed TF from blocking other interactions of its partners. The change in the specificity of the pairs may also explain why many sites that do not have a strong motif match are occupied by TFs *in vivo*. Importantly, the common interactions between the TFs resulted in a huge expansion of the lexicon of DNA words, suggesting that the gene regulatory code is much more complex than previously thought. The information about specific sequences bound by TF pairs is invaluable for understanding the ground rules for combinatorial gene regulation, which is central in tissue-specific responses to signals, and in cell-fate determination. Only a small subset of TF–TF pairs were investigated, so the role of cooperative TF–TF complexes is likely to be greater than our initial findings suggest. We plan to test a larger number of TF–TF combinations, and to study the role of DNA in interactions between TFs and other proteins involved in gene regulation, such as nucleosomes and histone-modifying enzymes. ■



The full version of this article is at Nature's website at go.nature.com/28sd9n9

GRAPHICAL ABSTRACT



Investigating TF–TF interactions in the presence of DNA by using CAP-SELEX. In CAP-SELEX, the two TFs are mixed with a complex pool of DNA molecules, followed by isolation of the complexes in which the two TFs bind the same DNA. The DNA in the complexes is then sequenced and data are analysed to detect cooperative binding either through detection of a composite site (for example, GCM1:ELK1, middle), which is often different (black boxes) from the individual sites, or by determining whether the individual characteristic 5-base-pair (bp) subsequences (blue boxes) of the two TFs frequently occur at a particular orientation and distance from each other (for example, MEIS1:DLX3, bottom). The heat map below the logos shows the frequency of each spacing and orientation; note that the 4-bp gap in the orientation shown is strongly preferred. The GCM1:ELK1 structure shown is a schematic presentation based on aligning the DNA sequence of the individual structures into a B-DNA model, whereas the MEIS1:DLX3 structure was solved in this study by using X-ray crystallography.

FURTHER READING

- Levine, M. Transcriptional enhancers in animal development and evolution. *Curr. Biol.* **20**, R754–R763 (2010).
This review broadly covers the role of transcriptional enhancers in development.
- Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
This study shows that the expression of combinations of TFs can reprogram cell fate.
- Meireles-Filho, A. C. A. *et al.* *cis*-Regulatory requirements for tissue-specific programs of the circadian clock. *Curr. Biol.* **24**, 1–10 (2014).
Combinatorial action with tissue-specific TFs explains cell-type-specific responses to the common TF regulated by the circadian clock.
- Yan, J. *et al.* Transcription factor binding in human cells occurs in dense clusters formed around cohesin anchor sites. *Cell* **154**, 801–813 (2013).
TF binding in human cells is located in dense clusters, and clustering is also observed for TFs that are not functionally related.
- Yip, K. Y. *et al.* Classification of human genomic regions based on experimentally determined binding sites of more than 100 transcription-related factors. *Genome Biol.* **13**, R48 (2012).
This study identifies regions in the genome bound by many TFs.
- Newman, J. R. S. & Keating, A. E. Comprehensive identification of human bZIP interactions with coiled-coil arrays. *Science* **300**, 2097–2101 (2003).
This paper analyses protein–protein interactions between bZIP family TFs.
- Gajiwala, K. S. *et al.* Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. *Nature* **403**, 916–921 (2000).
This structural study reveals the molecular mechanism of DNA binding by RFX1.
- Kim, S. *et al.* Probing allostery through DNA. *Science* **339**, 816–819 (2013).
The effects on DNA vibrational modes lead to cooperativity between distant proteins.
- Panne, D., Maniatis, T. & Harrison, S. C. An atomic model of the interferon-beta enhanceosome. *Cell* **129**, 1111–1123 (2007).
This study describes a model for the interferon enhanceosome.

RELATED CONTENT

Transcriptional regulators form diverse groups with context-dependent regulatory functions

G. Stampfel *et al.*
go.nature.com/28qksqw

Gene regulation: Landscape and mechanisms of transcription factor cooperativity

J. Osório
go.nature.com/28pdkpe