1-1-2008

Multiple YY1 and CTCF binding sites in imprinting control regions

Joomyeong Kim
Louisiana State University

Recommended Citation
Kim, J. (2008). Multiple YY1 and CTCF binding sites in imprinting control regions. *Epigenetics, 3* (3), 115-118. https://doi.org/10.4161/epi.3.3.6176

Epigenetics

ISSN: 1559-2294 (Print) 1559-2308 (Online) Journal homepage: https://www.tandfonline.com/loi/kepi20

Multiple YY1 and CTCF binding sites in imprinting control regions

Joomyeong Kim

To cite this article: Joomyeong Kim (2008) Multiple YY1 and CTCF binding sites in imprinting control regions, Epigenetics, 3:3, 115-118, DOI: 10.4161/epi.3.3.6176

To link to this article: https://doi.org/10.4161/epi.3.3.6176

[Epigenetics 3:3, 115-118; May/June 2008]; ©2008 Landes Bioscience

Point of View

Multiple YY1 and CTCF binding sites in imprinting control regions

Joomyeong Kim

Department of Biological Sciences; Louisiana State University; Baton Rouge, Louisiana USA

Key words: genomic imprinting, imprinting control regions, YY1, CTCF, DNA-binding sites

Known imprinting control regions (ICRs) contain unusual tandem arrays of DNA-binding sites for transcription factors, including YY1 for the *Peg3*, *Gnas* and *Xist/Tsix* domains and CTCF for the *H19/Igf2* domain. These multiple DNA-binding sites are known to be the only functionally shared and evolutionarily selected feature among these ICRs. However, it is not well understood why the imprinting control regions tend to maintain a high density of a particular transcription factor-binding site. We hypothesize that the multiplicity associated with the YY1 and CTCF binding sites may be designed for attracting and maintaining the relatively high levels of YY1 and CTCF proteins or for covering the relatively large genomic sizes of the associated ICRs. This idea remains to be tested in the near future, but it is one of the most likely explanations for all those unusual features that are associated with the functionally critical regions (ICRs) of genomic imprinting.

In mammals, a small subset of autosomal genes are not functionally equal between two parental alleles due to genomic imprinting, by which one allele is epigenetically repressed based on its parental origin. Imprinted genes tend to be clustered in specific regions of chromosomes, and the imprinting of a group of genes located in a given chromosomal domain is usually regulated through small genomic regions, termed imprinting control regions (ICRs). Known ICRs are usually CpG islands that are located close to the promoters of imprinted genes. The CpG sites within these ICRs are differentially methylated between two alleles, and this allele-specific DNA methylation of the ICRs is inherited as a gametic signal from the previous generation's germ cells. Mouse mutagenesis and human patients' studies have indicated that the ICRs play critical roles in (1) establishing and maintaining allele-specific DNA methylation status and (2) transcriptional control of neighbor genes.[1-3]

According to the information derived from the genomic sequences of several mammals, the sequence features associated with the ICRs appear to be very enigmatic. They do not show high levels of sequence conservation between different species, which is a hallmark of all the known regions involved in transcriptional regulation, e.g., enhancers and promoters. Instead, they often exhibit unusual species-specific tandem repeat sequence structures. However, careful examinations of these repeat sequences have identified small core sequences that are conserved between different species, which appear to be DNA-binding sites for several transcription factors. One well-known case is the binding sites for CTCF (CCCTC-binding factor) that are located in the ICR of *H19/Igf2*, and the other is the binding sites for YY1 (Yin-Yang 1) located in the ICRs of *Peg3*, *Nespas*, *Tsix* and *Xist*.[4-7] CTCF binding sites are also shown to be located within the following imprinted genes, *Kcnq1*, *Gtl2* and *Rasgrf1*.[8-10] Several functional studies demonstrated that these binding sites are very crucial for maintaining the DNA methylation status of the associated ICRs. Either complete removal or point mutations of CTCF binding sites resulted in hypermethylation on the maternal allele of H19-ICR, suggesting a protector role of CTCF against DNA methylation.[11-14] On the other hand, recent studies that reduce the cellular levels of YY1 have hinted a similar role for YY1 in the *Peg3*, *Nespas* and *Xist* domains. siRNA knockdown-based reduction of YY1 levels caused target-specific changes in the methylation levels of these ICRs. In the cell line-based experiments,[15] YY1 knockdown caused hypomethylation, suggesting YY1 role in maintaining somatic maternal allele DNA methylation. In contrast, in vivo YY1 knockdown caused both hyper and hypomethylation, suggesting that YY1 may be involved in both protecting and maintaining the somatic DNA methylation status of Peg3-ICR.[16] Besides the observed role in DNA methylation, both CTCF and YY1 are also involved in the transcriptional control of the associated imprinted domains. For the *H19/Igf2* domain, CTCF acts as an enhancer-blocking insulator coordinating the reciprocal, allele-specific expression of *H19* and *Igf2*. CTCF is also involved in the initiation of H19 promoter transcription[14] and chromosome loop formation for the *H19/Igf2* domain.[17] For the YY1-associated imprinted domains, YY1 appears to be either an orientation-dependent activator or repressor for the transcription of neighboring genes. In sum, the results described above clearly indicate that the DNA-binding sites for both YY1 and CTCF are the most functionally significant *cis*-elements relevant to the predicted roles of the ICRs.

The DNA-binding sites for YY1 and CTCF located within the ICRs of *Peg3* and *H19* show very unusual features that have never been described in other normal genes (Fig. 1). Within a 3 kb genomic interval of Peg3's 1st intron, 6 to 14 YY1 binding sites are found in each species of mammals. Half of these DNA binding sites are perfectly matched with the consensus sequence of YY1 binding motifs, while the remaining half shows one or two base differences

Correspondence to: Joomyeong Kim; Department of Biological Sciences; 202 Life Sciences Building; South Campus Rd.; Louisiana State University; Baton Rouge, 70803 Louisiana USA; Tel.: 225.578.7692; Fax: 225.578.2597; Email: jkim@lsu.edu

Submitted: 01/30/08; Accepted: 04/18/08

Previously published online as an *Epigenetics* E-publication:
http://www.landesbioscience.com/journals/epigenetics/article/6176

from the consensus. The surrounding regions of each YY1 binding site maintain high levels of sequence similarity within species, but not between different species. This suggests that the YY1 binding sites have been duplicated independently in each lineage of mammals as part of tandem repeats. Consequently, the spacing and positions of YY1 binding sites are not conserved between different species. Interestingly, however, the orientations of the YY1 binding sites are all identical, and also conserved between different mammals. Very similar patterns are also observed in the CTCF binding sites located within the H19-ICR (Fig. 1B). Within a 2 to 5 kb genomic interval, 4 to 10 CTCF binding sites are found in each species of mammals and the relative positions and spacing of individual CTCF binding sites are not conserved. However, the orientations of the CTCF binding sites are all identical and also conserved between different mammals.

The actual sequences of the DNA-binding motifs for YY1 and CTCF are also very unusual (Fig. 1). They appear to have maintained very high levels of sequence homogeneity during evolution. This is very unusual given the fact that most DNA-binding transcription factors can bind to a set of different sequences, which are usually represented by a consensus sequence with variable bases in allowable positions. In fact, the consensus motif for YY1 is represented by a 9 bp-long sequence (CGCCATnTT; n means any bases at that position), while the shorter core motif (GCCAT) is absolutely required for the binding to YY1 protein. Nonetheless, all of the YY1 motifs found in the Peg3-ICR are one type (GGCGCCATCTT) without any base variations, and are even slightly longer than the known consensus motif of YY1. According to our unpublished data, this longer motif is indeed a previously unrecognized motif of YY1, but one that displays much higher levels of DNA-binding affinity to the YY1 protein than the shorter consensus motif. A similar situation is also observed in the CTCF binding sites of the H19-ICR (Fig. 1B). Although CTCF is known to bind to a set of very loose and degenerate sequences,[18,19] the CTCF binding sites found within the H19-ICR show high levels of sequence homogeneity between individual binding sites and also between different species. It is unknown whether the observed sequence homogeneity also results in higher levels of binding affinity to the CTCF sites in the H19-ICR. Overall, the DNA-binding sites of both YY1 and CTCF found in the ICRs exhibit the same very unusual features: high densities of multiple binding sites with an identical orientation.

Both YY1 and CTCF proteins are evolutionarily well conserved: the homologues of both proteins are found in organisms ranging from flying insects to mammals.[20-23] Both proteins are known to play critical roles for the various aspects of transcription and epigenetic modification. YY1 is known to control about 10% of mammalian genes, and is also shown to interact with many key proteins involved in transcriptional control. Notably, YY1 is a protein targeting the Polycomb complex, a repression machinery for flying insets and vertebrates. Consistently, YY1 binding sites are frequently found in many heterochromatic regions of genomes, including pericentromeric repeats, L1, IAP, ERV and even one family of SINEs (Alu in human). Similarly, CTCF is also involved in the various aspects of transcription and chromatin formation. CTCF was initially identified as a repressor protein binding to the promoters of cMyc and lysozyme, and later as an activator for other genes. More importantly, as described earlier, CTCF has been recognized as the

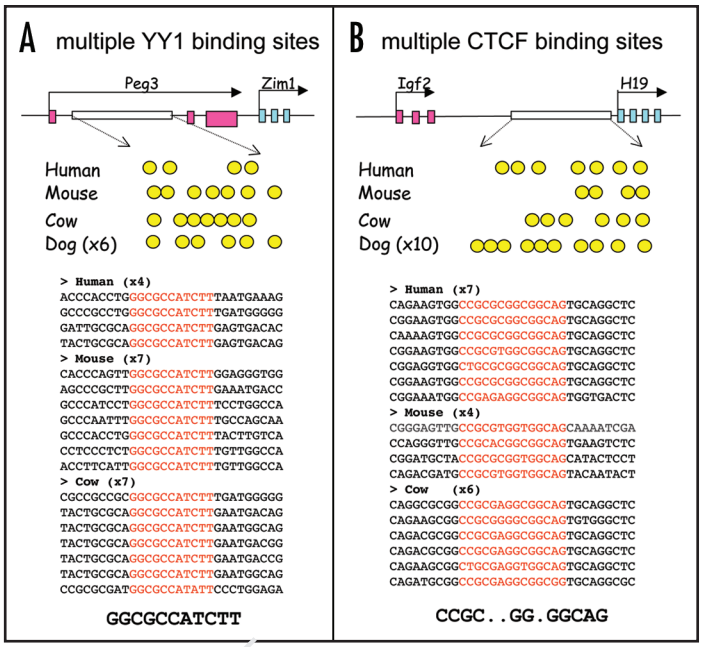

Figure 1. Multiple DNA-binding sites for YY1 (A) and CTCF (B) found in the ICRs of *Peg3* and *H19*, respectively. Multiple DNA-binding sites for YY1 and CTCF are indicated by yellow circles. The positions and spacing of the DNA-binding sites for YY1 and CTCF are not conserved between different mammals. In contrast, the orientations of individual binding sites are all identical and also conserved between different species. Sequence alignments of the DNA-binding sites and surrounding regions are shown in the bottom.

first insulator protein identified from vertebrates. Recent studies further highlighted that this protein may be a global organizer for the chromatin structure of mammalian genomes.[24] Also, CTCF is involved in asynchronous replication,[25] inter-chromosomal interactions,[26] and organizing sister chromatids as a component for the mammalian cohesin complex.[27,28] Consistent with the diverse roles played by YY1 and CTCF, the expression patterns and levels of these two proteins are ubiquitous and abundant in vertebrates. Genetic studies with mutant mice further indicated that the dosage of these proteins are very critical for the survival of animals,[16,29,30] suggesting that the cellular levels of these two proteins may be tightly controlled for their appropriate distribution to various cellular functions.

Both YY1 and CTCF proteins are regarded as general transcription factors based on their various and diverse roles, but not as special proteins designed for genomic imprinting, a mammal-specific mechanism controlling the dosage of less than 200 genes. Then, how could these general transcription factors become key proteins for this lineage-specific mechanism? This may be facilitated by the unusual multiple transcription factor-binding sites located within imprinting control regions (ICRs). Since tandem arrays of transcription factor-binding sites are mainly found within the ICRs, these may be required for the functions of the ICRs. The following are potential scenarios how these multiple DNA-binding sites are involved in the ICR functions. First, multiple binding sites for YY1 and CTCF may be required simply to cover the relatively large genomic sizes of the ICRs, 1 to 4 kb in length, which are considered to be too large to be regulatory regions given the much smaller sizes of other regulatory regions, such as promoters and enhancers that are 100 to 500 bp in length. In particular, one or two binding sites may not be

sufficient to establish and maintain DNA methylation on these ICRs. This may be one reason why multiple binding sites for YY1 and CTCF have been selected within the ICRs, simply to cover large genomic distances. Second, the special functions of ICRs may require much higher levels of the YY1 and CTCF proteins than other genes that are controlled by these two proteins. The clustered binding sites may function as 'genomic trappers' that recruit and secure a sufficient amount of the YY1 and CTCF proteins around the ICRs. ICRs with multiple high-affinity binding sites for YY1 and CTCF should out compete to attract greater amounts of proteins and to confine them around their genomic regions for a longer time than any other regions. Consequently, the ICRs with multiple binding sites of YY1 or CTCF would be able to maintain relatively high concentrations of these proteins around their genomic regions. These high levels of the YY1 or CTCF protein may be critical for the functions of the ICRs. If this prediction is true, the functions of the ICRs should be very susceptible to changes in the levels of these proteins. This has been somewhat confirmed through knockdown experiments targeting YY1 and CTCF.[13,16] In short, the reduced levels of either YY1 or CTCF protein have a significant impact on the DNA methylation levels of the associated ICRs. These knockdown effects are also shown to be target-specific and do not affect other imprinted domains. The two possibilities described above are not mutually exclusive, and thus these possibilities might be two separate functional constraints that have formed the multiple binding sites of YY1 and CTCF in the ICRs.

There are a number of intriguing features that require further attention besides the multiplicity associated with the DNA-binding sites of YY1 and CTCF. For example, all those binding sites of YY1 and CTCF are in a same direction and also maintain unusually high levels of sequence homogeneity. The orientation of DNA-binding sites may be related to the transcriptional direction of nearby genes, such as *Peg3* and *H19*. There might be some extra constraints for these binding sites to maintain a particular direction relative to the transcriptional direction of either *Peg3* or *H19*. Another puzzle is the evolutionary selection of one particular type of DNA-binding motif for either YY1 or CTCF within the ICRs. This may not be simply designed to increase their binding strength since other motifs can have similar strength as those shown in the ICRs. Besides YY1 and CTCF, more transcription factors are predicted to be involved in genomic imprinting based on the fact that the ICRs of other imprinted domains also show similar tandem repeat sequence structures with evolutionary conservation.[31,32] We need to first test if some of these repeats also contain DNA-binding sites for other transcription factors, but, if that is the case, it will be very interesting to characterize what functional constraints have maintained multiple DNA-binding sites for these imprinted domains. Based on many similarities shared between different ICRs, it is most likely that the multiplicity associated with these unknown transcription factors along with YY1 and CTCF may have similar functional roles for genomic imprinting.

Acknowledgements

This work was supported by NIH (R01-GM66225).

References

1. Brannan CI and Bartolomei MS. Mechanisms of genomic imprinting. Curr Opin Genet Dev 1999; 9:164-70.
2. Spahn L and Barlow DP. An ICE pattern crystallizes. Nat Genet 2003; 35:11-2.
3. Lewis A and Reik W. How imprinting centres work. Cytogenet Genomic Res 2006; 113:81-9.
4. Bell AC and Felsenfeld G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature 2000; 405:482-5.
5. Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM and Tilghman SM. CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature 2000; 405:486-9.
6. Kim J, Kollhoff A, Bergmann A and Stubbs L. Methylation-sensitive binding of transcription factor YY1 to an insulator sequence within the paternally expressed imprinted gene, Peg3. Hum Mol Genet 2003; 12:233-45.
7. Kim JD, Hinz AK, Bergmann A, Huang JM, Ovcharenko I, Stubbs L and Kim J. Identification of clustered YY1 binding sites in imprinting control regions. Genome Res 2006; 16:901-11.
8. Fitzpatrick GV, Pugacheva EM, Shin JY, Abdullaev Z, Yang Y, Khatod K, Lobanenkov VV and Higgins MJ. Allele-specific binding of CTCF to the multipartite imprinting control region KvDMR1. Mol Cell Biol 2007; 27:2636-47.
9. Yoon B, Herman H, Hu B, Park YJ, Lindroth A, Bell A, West AG, Chang Y, Stablewski A, Piel JC, Loukinov DI, Lobanenkov VV and Soloway PD. Rasgrf1 imprinting is regulated by a CTCF-dependent methylation-sensitive enhancer blocker. Mol Cell Biol 2005; 25:11184-90.
10. Charlier C, Segers K, Wagenaar D, Karim L, Berghmans S, Jaillon O, Shay T, Weissenbach J, Cockett N, Gyapay G and Georges M. Human-ovine comparative sequencing of a 250 kb imprinted domain encompassing the callipyge (clpg) locus and identification of six imprinted transcripts: DLK1, DAT, GTL2, PEG11, antiPEG11 and MEG8. Genome Res 2001; 11:850-62.
11. Schoenherr CJ, Levorse JM and Tilghman SM. CTCF maintains differential methylation at the Igf2/H19 locus. Nat Genet 2003; 33:66-9.
12. Engel N, West AG, Felsenfeld G and Bartolomei MS. Antagonism between DNA hypermethylation and enhancer-blocking activity at the H19 DMD is uncovered by CpG mutations. Nat Genet 2004; 36:883-8.
13. Fedoriw AM, Stein P, Svoboda P, Schultz RM and Bartolomei MS. Transgenic RNAi reveals essential function for CTCF in H19 gene imprinting. Science 2004; 303:238-40.
14. Engel N, Thorvaldsen JL and Bartolomei MS. CTCF binding sites promote transcription initiation and prevent DNA methylation on the maternal allele at the imprinted H19/Igf2 locus. Hum Mol Genet 2006; 15:2945-54.
15. Kim JD, Hinz AK, Choo JH, Stubbs L and Kim J. YY1 as a controlling factor for the Peg3 and Gnas imprinted domains. Genomics 2007; 89:262-9.
16. Kim J and Kim JD. In vivo YY1-knockdown effects on genomic imprinting. Hum Mol Genet 2008; 17:391-401.
17. Kurukuti S, Tiwari VK, Tavoosidana G, Pugacheva E, Murrell A, Zhao Z, Lobanenkov V, Reik W and Ohlsson R. CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. Proc Natl Acad Sci USA 2006; 103:10684-9.
18. Filippova GN, Fagerlie S, Klenova EM, Myers C, Dehner Y, Goodwin G, Neiman PE, Collins SJ and Lobanenkov VV. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. Mol Cell Biol 1996; 16:2802-13.
19. Renda M, Baglivo I, Burgess Beusse B, Esposito S, Fattorusso R, Felsenfeld G and Pedone PV. Critical DNA binding interactions of the insulator protein CTCF: a small number of zinc fingers mediate strong binding, and a single finger-DNA interaction controls binding at imprinted loci. J Biol Chem 2007; 282:33336-45.
20. Shi Y, Lee JS and Galvin KM. Everything you have ever wanted to know about Yin Yang 1. Biochim Biophys Acta 1997; 1332:49-66.
21. Thomas MJ and Seto E. Unlocking the mechanisms of transcription factor YY1: are chromatin modifying enzymes the key? Gene 1999; 236:197-208.
22. Gordon S, Akopyan G, Garban H and Bonavida B. Transcription factor YY1: structure, function, and therapeutic implications in cancer biology. Oncogene 2006; 25:1125-42.
23. Filippova GN. Genetics and epigenetics of the multifunctional protein CTCF. Curr Top Dev Biol 2008; 80:337-60.
24. Wallace JA and Felsenfeld G. We gather together: insulators and genome organization. Curr Opin Genet Dev 2007; 17:400-7.
25. Bergström R, Whitehead J, Kurukuti S and Ohlsson R. CTCF regulates asynchronous replication of the imprinted H19/Igf2 domain. Cell Cycle 2007; 6:450-4.
26. Ling JQ, Li T, Hu JF, Vu TH, Chen HL, Qiu XW, Cherry AM and Hoffman AR. CTCF mediates inter-chromosomal colocalization between Igf2/H19 and Wsb1/Nf1. Science 2006; 312:269-72.
27. Parelho V, Hadjur S, Spivakov M, Leleu M, Sauer S, Gregson HC, Jarmuz A, Canzonetta C, Webster Z, Nesterova T, Cobb BS, Yokomori K, Dillon N, Aragon L, Fisher AG and Merkenschlager M. Cohesins functionally associate with CTCF on mammalian chromosome arms. Cell 2008; 132:422-33.

28. Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, Tsutsumi S, Nagae G, Ishihara K, Mishiro T, Yahata K, Imamoto F, Aburatani H, Nakao M, Imamoto N, Maeshima K, Shirahige K and Peters JM. Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature 2008; 451:796-801.
29. Donohoe ME, Zhang X, McGinnis L, Biggers J, Li E and Shi Y. Targeted disruption of mouse Yin Yang 1 transcription factor results in peri-implantation lethality. Mol Cell Biol 1999; 19:7237-44.
30. Affar el B, Gay F, Shi Y, Liu H, Huarte M, Wu S, Collins T and Li E. Essential dosage-dependent functions of the transcription factor yin yang 1 in late embryonic development and cell cycle progression. Mol Cell Biol 2006; 26:3565-81.
31. Hutter B, Helms V and Paulsen M. Tandem repeats in the CpG islands of imprinted genes. Genomics 2006; 88:323-32.
32. Paoloni Giacobino A, D'Aiuto L, Cirio MC, Reinhart B and Chaillet JR. Conserved features of imprinted differentially methylated domains. Gene 2007; 399:33-45.