7236–7253 *Nucleic Acids Research, 2023, Vol. 51, No. 14* *Published online 19 June 2023*
https://doi.org/10.1093/nar/gkad516

Five nucleotides found in RCTG motifs are essential for post-fertilization methylation imprinting of the *H19* ICR in YAC transgenic mice

Hitomi Matsuzaki[1], Takuya Takahashi[2], Daichi Kuramochi[2], Katsuhiko Hirakawa[2] and Keiji Tanimoto[1,*]

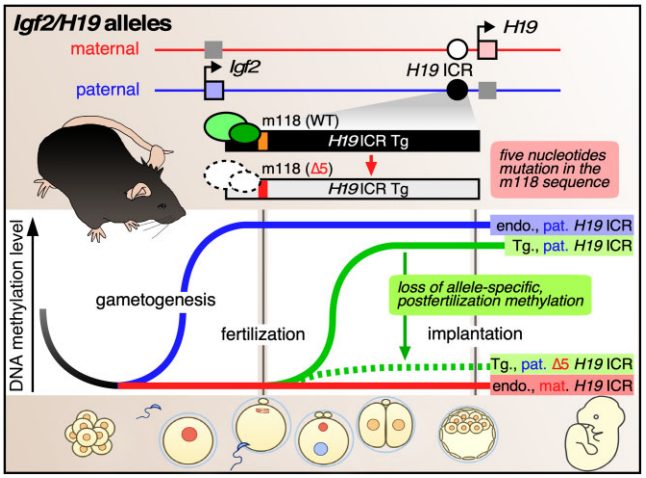
[1]Faculty of Life and Environmental Sciences, Life Science Center for Survival Dynamics, Tsukuba Advanced Research Alliance (TARA), University of Tsukuba, Tsukuba, Ibaraki 305-8577, Japan and [2]Graduate school of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki 305-8577, Japan

Received October 18, 2022; Editorial Decision May 20, 2023; Accepted June 02, 2023

ABSTRACT

Genomic imprinting at the mouse *Igf2*/*H19* locus is controlled by the *H19* ICR, within which paternal allele-specific DNA methylation originating in sperm is maintained throughout development in offspring. We previously found that a 2.9 kb transgenic *H19* ICR fragment in mice can be methylated *de novo* after fertilization only when paternally inherited, despite its unmethylated state in sperm. When the 118 bp sequence responsible for this methylation in transgenic mice was deleted from the endogenous *H19* ICR, the methylation level of its paternal allele was significantly reduced after fertilization, suggesting the activity involving this 118 bp sequence is required for methylation maintenance at the endogenous locus. Here, we determined protein binding to the 118 bp sequence using an *in vitro* binding assay and inferred the binding motif to be RCTG by using a series of mutant competitors. Furthermore, we generated *H19* ICR transgenic mice with a 5-bp substitution mutation that disrupts the RCTG motifs within the 118 bp sequence, and observed loss of methylation from the paternally inherited transgene. These results indicate that imprinted methylation of the *H19* ICR established *de novo* during the post-fertilization period involves binding of specific factors to distinct sequence motifs within the 118 bp sequence.

GRAPHICAL ABSTRACT



INTRODUCTION

Genomic imprinting is an epigenetic phenomenon in mammals, in which a subset of genes is expressed from only one of two alleles, depending on parental origin. Since the first imprinted genes were discovered in mice (1–3), approximately 150 genes have been identified. Of these 150, many are conserved among mammalian species and essential for both fetal and placental growth and for development (4,5). Their misexpression can therefore cause epigenetic disorders in mammals, including humans. Hence, a complete understanding of the molecular mechanisms underlying this mono-allelic gene expression mechanism is crucial (6).

Imprinted gene expression is regulated by chemical modifications of chromatin. The best known modification is DNA methylation. In many cases, imprinted genes cluster to form imprinted loci that are frequently accompanied by a *cis*-regulatory sequence called the Imprinting Control Region (ICR) (6,7). The ICR is subject to allele-specific DNA

*To whom correspondence should be addressed. Tel: +81 298 53 7300; Fax: +81 298 53 6070; Email: keiji@tara.tsukuba.ac.jp

methylation to regulate imprinting throughout the locus. Since such imprinted methylation is acquired in the oocyte or sperm, the ICRs are also called germline differentially methylated regions (gDMRs). In the canonical form of imprinting that is regulated by DNA methylation, most ICRs are methylated in the oocyte and coincide with the promoter CpG islands of the imprinted genes. In contrast, three ICRs, including the *H19* ICR at the *Igf2*/*H19* locus, are DNA-methylated in sperm, at the intergenic region (8,9).

A non-canonical imprinting mechanism was recently reported. Genomic loci that undergo this type of imprinting show parental bias in gene expression after fertilization even though they do not show DNA methylation bias in the germline. Moreover, histone modification by H3K27me3 in the oocyte is reportedly the underlying mechanism (10). Gene expression from the allele of maternal origin carrying this modification is suppressed in the pre-implantation embryo. While this modification is gradually lost in the embryo, it is converted to DNA methylation modification in the placenta to ensure genomic imprinting thereafter (11,12).

The *H19* ICR of the *Igf2*/*H19* gene locus has been shown to undergo canonical imprinting. The *H19* ICR DNA is methylated in sperm. After fertilization, only the paternal allele remains methylated throughout development (13–16). Allele-specific methylation of the *H19* ICR regulates monoallelic gene expression of both *Igf2* and *H19* through multiple mechanisms. Transcription factor CTCF (CCCTC-Binding Factor) binds to the unmethylated *H19* ICR on the maternally inherited allele to form an insulator that blocks *Igf2* gene activation by the 3′ enhancer shared by both genes (17–20). Unmethylated *H19* ICR is also responsible for full activation of *H19* gene in a direction-dependent manner (21). On the other hand, CTCF cannot bind to the methylated *H19* ICR in the paternal allele, so the insulator does not form and *Igf2* expression is fully activated. At the same time, the methylated *H19* ICR induces methylation of the downstream *H19* promoter and represses its expression (21–23).

In order to elucidate the mechanism underlying imprinted DNA methylation of the *H19* ICR, we previously tested the activity of the 2.9 kb *H19* ICR sequence *in vivo* (24). We inserted this DNA fragment at the human β-globin gene locus, which normally does not undergo genomic imprinting, on a yeast artificial chromosome (YAC). We generated transgenic mice (TgM) using this construct and found that in somatic cells, the transgenic *H19* ICR sequence was DNA-methylated only when paternally inherited, indicating that the *H19* ICR has intrinsic activity for establishing imprinted DNA methylation. Surprisingly, however, the *H19* ICR transgene was unmethylated in sperm. Instead, this transgene acquired DNA methylation in the early embryo after fertilization in a paternal-allele specific manner (25). In other words, the *H19* ICR transgene should have inherited epigenetic modification bias, other than DNA methylation, from the germlines. Based on this bias, the alleles must be distinguished immediately after fertilization and *de novo* DNA-methylated in a paternal allele-specific manner. This result also suggests that the *H19* ICR, which is widely believed to undergo canonical imprinting, may also possess characteristics similar to those of sequences that undergo non-canonical imprinting. However, unlike the non-canonical imprinting example described earlier, the germline-derived epigenetic signature is apparently converted to DNA methylation during the pre-implantation period at the transgenic *H19* ICR (25).

We next investigated *cis*-regulatory sequences involved in post-fertilization imprinted methylation of the *H19* ICR (25–27). We found that a 118 bp sequence at the 5′ end of the 2.9 kb *H19* ICR is required for imprinting of transgenic *H19* ICR (28). While deleting the same sequence from the endogenous *H19* ICR did not affect its DNA methylation level in sperm, it reduced the DNA methylation level of the paternally inherited *H19* ICR immediately after fertilization and caused loss of imprinted expression of both *Igf2* and *H19* in embryos (25,27). Thus, in pre-implantation embryos, the 118 bp sequence may function in distinguishing the *H19* ICR from other genomic sequences and recognizing its parental origin, to introduce allele-specific DNA methylation, thereby protecting the sequence against non-allele specific demethylation activity. Recently, a subset of mouse genomic sites in the promoter region of the paternal allele was found to acquire DNA methylation in pre-implantation embryos, suggesting that allele-specific, post-fertilization methylation is not restricted to imprinted gene loci (29). Subsequently, we demonstrated that the 118 bp sequence is sufficient to confer paternal allele-specific DNA methylation to a lambda phage DNA-based artificial sequence during the post-fertilization period (28).

Within the 118 bp sequence, a *cis*-regulatory DNA motif or motifs are quite likely recognized by a specific *trans*-acting factor(s) that functions in either discriminating between the alleles, recruiting DNA methyltransferase, or both. The *trans*-acting factors Zfp57 and Zfp445 are reportedly involved in maintaining paternal allele-specific DNA methylation of the *H19* ICR as well as other ICRs. Zfp57 and Zfp445 are thought to bind to consensus recognition sequences within the ICRs (TGCCGC; six such sites are present in the mouse *H19* ICR) and recruit Trim28 (Kap1) and DNA methyltransferases (30–32). Although CpG motifs in the recognition sequence must be methylated for these factors to bind, neither consensus binding sequences nor CpG motifs are present within the 118 bp sequence (28). Moreover, the binding of such factors was not observed in our *in vitro* binding assay (27). Therefore, we predict that another regulatory factor(s) operates within the 118 bp sequence of the *H19* ICR and induces DNA methylation.

In the work described here, we sought to investigate the molecular mechanisms of DNA methylation that involve the 118 bp sequence. We searched for protein-binding sequence motifs using *in vitro* binding assays and identified five RCTG motifs as candidate binding sites for regulatory protein(s). In addition, we found that TR2/4 proteins bind direct repeat (DR) sequences present at the 3′ end of the 118 bp sequence. We proceeded to generate two kinds of mutant *H19* ICR (2.9 kb) fragments by introducing either a five-nucleotide substitution (Δ5) or a 38 bp deletion (LCb80) to disrupt the five RCTG motifs and the DR sequence within the 118 bp sequence, respectively. We used the mutant constructs to generate TgM. We found that imprinted DNA methylation was no longer established after fertilization in Δ5 TgM, while it was partially impaired on the paternally

inherited allele in LCb80 TgM. These results indicate that specific *cis* sequences and their binding factors are involved in post-fertilization imprinted methylation of the *H19* ICR.

MATERIALS AND METHODS

Cell culture, protein ablation by small interfering (si) RNAs, and western blot analysis

P19 cells were maintained in Dulbecco's modified Eagle's medium (DMEM, 08456–36; nacalai tesque) containing 10% fetal bovine serum (FBS) and penicillin-streptomycin.

The siRNA duplexes (ON-TARGET*plus* SMART pools) were purchased from Horizon Discovery and transfected into P19 cells using Lipofectamine RNAiMAX (Invitrogen). The sequences of the siRNA pools were as follows: non-targeting pool, UGGUUUACAUGUCGACUAA, UGGUUUACAUGUUGUGUGA, UGGUUUACAUGUUUUCUGA, and UGGUUUACAUGUUUUCCUA; mouse Nr2c1 (TR2) siRNA pool, GGUGCAGAGCUUACGCAAU, GGAUCCAAAGACUGCGUUA, CCGGAAAGGAGGAAGUCGU, and UCAAUAAGCACCACCGAAA; mouse Nr2c2 (TR4) siRNA pool, GAUCCUGGCUUCUCCGGAA, AUUUGACACCUUAGCGAAA, GAGAAGAUCUAUAUCCGGA, and CUGUACAGAGUGAACGGAA.

Nuclear extracts were separated by 10% SDS-PAGE. Separated proteins were transferred to a polyvinylidene difluoride membrane (Millipore) and blotted with antibodies against TR2 (M-85; cat. no. sc-9087; Santa Cruz Biotechnology), TR4 (M-76; cat. no. sc-9086; Santa Cruz Biotechnology), and PCNA (cat. no. 610665, BD Transduction Laboratories).

Electrophoretic mobility shift assay (EMSA)

Nuclear extracts were prepared from mouse P19 cells, ES cells, or mouse testes of adult males by using Nuclear Extract Kit (Active Motif) according to the manufacturer's instructions. Testes were homogenized in the ice-cold sucrose buffer [0.25 M sucrose, 1 mM EDTA, 3 mM imidazole, 0.1% (v/v) ethanol and Protease Inhibitor Cocktail (nacalai tesque), pH 7.2] and cells were pelleted by centrifugation (700g, 10 min, 4°C) and subjected to nuclear extraction. Nuclear extracts (7 μg, unless otherwise stated) were preincubated in the reaction mixture [PBS with 5 mM $MgCl_2$, 0.1 mM $ZnSO_4$, 1 mM DTT, 0.1% NP40, 10% glycerol and 1 μg of poly(dI-dC)] for 10 min at RT, with or without 20–200-fold molar excess of a specific double-stranded competitor DNA. For super-shift assays, 1 μg of above mentioned antibodies specific for either TR2 or TR4 (Santa Cruz Biotechnology) was included in the reaction mixture. Radiolabeled DNA probe (15000 cpm) was added and the incubation was continued for 25 min at RT. The incubation mixture was loaded on a 3.5 or 4% non-denaturing polyacrylamide gel in 0.5× TBE buffer, and electrophoresed at 4°C. The gels were dried and exposed to X-ray film. Probe and competitor sequences are indicated in the figures.

Generation of genetically modified mice

Preparation of the LCb 80 fragment. Two DNA fragments were PCR-generated using either the murine *H19* ICR DNA as a template and a set of primers: 5′del_fr-3A8G + B + X, 5′-GGATCTAGAGATCTGGATCCAAGCTTTCCTGCTCACTG-3′ (*Xba*I, *Bgl*II, *Bam*HI and *Hin*dIII sites underlined) and ICRcore-80–3A, 5′-CTGAATTCTGGTCACAGCACTGCTATGT-3′ (*Eco*RI), or the λ DNA as a template and a set of primers: Lambda-5S2, 5′-CTGAATTCtcgagcttaatattctat-3′ (*Eco*RI; λ sequences in lower case letters) and LS5, 5′-TATTCTCGAGACGCGTTTTGCTGCCACCACGCGGCAACtaggtgttttaactcgtg-3′ (*Xho*I, *Mlu*I and CTCF binding sites are underlined; λ sequences in lower case letters). Resultant fragments were digested with *Bgl*II/*Eco*RI and *Eco*RI/*Mlu*I, respectively, linked together at their *Eco*RI ends to generate 5′-end of the LCb80 fragment. Preparation of λ+CTCF + b (LCb) sequence was described elsewhere (33). The LCb fragment, released by *Bam*HI digestion was blunt-ended and ligated with *Bgl*II linker (pCAGATCTG). 3′ segment of this fragment, carrying CTCF sites 2 to 4, was recovered by *Mlu*I/*Bgl*II digestion and ligated with the 5′-end of the LCb80 fragment (*Bgl*II-*Mlu*I fragments, described above) at their *Mlu*I ends. The LCb80 fragment was released by *Bgl*II digestion.

Preparation of the H19 *ICR* Δ*5 fragment.* In order to introduce five point-mutations (Δ5) into the *H19* ICR sequences, mutant fragment was PCR-generated by using following set of primers: Δ5–5S, 5′-TTTAGCCTGACCAAGGAAGCTTTCCTGCTCA***a***TGTCCATTCAATGCA***t***TCAAAAGTG***a***TGTGACTATACAGGAGGAACATAGCA***t***TG***a***TGTGACCATAC-3′ (*Eco*NI site is underlined and mutated nucleotides are bold-italicized) and ORI-5S1, 5′-TACCAGCCTAGAAAATGCATGTGT-3'. After *Eco*NI/*Bsu*36I digestion, the fragment was replaced with the corresponding portion of the *H19* ICR [nucleotides 1175–1727 (AF049091.1; GenBank)]. The *H19* ICR Δ5 fragment was released by *Bam*HI digestion.

Yeast targeting vectors and homologous recombination in yeast. The co-placement target vector, pHS1/loxP-5171-B-2272–5171-G-2272 (pCop5B25G2), carrying a human β-globin HS1 fragment [nucleotides 13299–14250 (HUMHBB; GenBank)], in which 5′-loxP5171-*Bam*HI-loxP2272-loxP5171-*Bgl*II-loxP2272-3′ sequences are introduced into the *Hin*dIII site [at nucleotide 13769 in HUMHBB], was described elsewhere (33).

The Δ5 was inserted into *Bam*HI site of pCop5B25G2 to generate pCop5[Δ5]25G2. The resultant plasmid was digested (partially, in the case of pCop5[Δ5]25G2) with *Bgl*II and ligated with another fragment LCb80 to generate pCop5[Δ5]25[LCb80]2. In each cloning step, the correctness of DNA construction was confirmed by DNA sequencing.

The targeting vector was linearized with *Spe*I [at nucleotide 13670 in HUMHBB] and used to mutagenize the human β-globin YAC (A201F4.3) (34). Successful homologous recombination in yeast was confirmed by Southern blot analyses with several combinations of restriction enzymes and probes.

Generation of YAC-TgM. Purified YAC DNA was microinjected into fertilized mouse eggs from C57BL/6J (Charles River) mice. Tail DNA from founder offspring was screened first by PCR, followed by Southern blotting. Structural analysis of the YAC transgene was performed as described elsewhere (34,35). Zp3-cre gene TgM (Jackson Laboratory) (36) were mated with parental YAC-TgM lines to generate sublines (*i.e.* each carrying one of the test fragments). Successful cre-loxP recombination was confirmed by Southern blotting.

Animal experiments were performed in a humane manner and approved by the Institutional Animal Experiment Commit- tee of the University of Tsukuba. Experiments were conducted in accordance with the Regulation of Animal Experiments of the University of Tsukuba and the Fundamental Guidelines for Proper Conduct of Animal Experiments and Related Activities in Academic Research Institutions under the jurisdiction of the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan.

Deletion of 36-bp sequence by genome-editing in mouse. Two sets of oligonucleotides were annealed and inserted at the *Bbs*I site of the pX330 (plasmid #42230; Addgene) (37) to generate Cas9/sgRNA expression vectors. For the 5′ border: 5′-caccGTAGCAGTGCTGTGACCATAC-3′ and 5′-aaacGTATGGTCACAGCACTGCTAC-3′; and for the 3′ border: 5′-caccGAACACACTTACATGGCACCA-3′ and 5′-aaacTGGTGCCATGTAAGTGTGTTC-3′ (overhanging nucleotides are shown in lowercase letters). The plasmids were microinjected into the pronuclei of fertilized eggs of ICR/β-globin TgM (24). Tail DNA from founder offspring was screened by PCR and sequencing for desired recombination event. The founders were then crossed with wild-type mice to establish TgM lines, in which the transgenic *H19* ICR was mutated, while the endogenous *H19* ICR was intact.

Preparation of embryos

Female mice were super-ovulated via injection of pregnant mare serum gonadotropin, followed by human chorionic gonadotropin (hCG) (47–48 h interval). Fertilized one-cell zygotes were collected from oviducts of mated females 24 h after hCG injection, and cumulus cells were removed by hyaluronidase treatment. Two-cell embryos were flushed from oviducts at 44h after hCG injection.

DNA methylation analysis by southern blotting

Genomic DNA extracted from tail tips of ∼1-week-old animals or testes of adult males was first digested by *Bam*HI, and then subjected to the methylation-sensitive enzyme *Bst*UI or *Hha*I. Following size separation in agarose gels, Southern blots were hybridized with α-^{32}P-labeled probes and subjected to X-ray film autoradiography.

DNA methylation analysis by bisulfite sequencing

Pre-implantation embryos were embedded in agarose beads and treated with sodium bisulfite as described previously (38). Genomic DNA extracted from tail tips or adult male sperm was treated with sodium bisulfite using the EZ DNA Methylation Kit (Zymo Research). Tail tip and sperm DNA was digested with XbaI prior to the treatment. Subregions of the transgenes were amplified by nested PCR. The PCR products were subcloned into the pGEM-T Easy vector (Promega) for sequencing analyses. PCR primers are listed in Tables 1 and 2.

Table 1. Primer sets for bisulfite sequencing analysis

Regions analyzed	Allele	PCR round	5′ primer	3′ primer
I	Tg-Δ5	1st	LCR-MA-5S1	ICR-MA-3A15
	Tg-Δ36	2nd	ICR-MA-5S4	ICR-MA-3A14
II	Tg-WT	1st	ICR-MA-5S13	BGLB-MA-3A2
	Tg-Δ5	2nd	ICR-MA-5S13	ICR-MA-3A26
II′	Tg-Δ36	1st	ICR-MA-5S13	BGLB-MA-3A2
		2nd	ICR-MA-5S13	ICR-MA-3A26
III	Tg-LCb80	1st	lambda-MA-5S4	lambda-MA-3A2
		2nd	lambda-MA-5S1	lambda-MA-3A3
IV	Tg-LCb80	1st	lambda-MA-5S5	lambda-MA-3A7
		2nd	lambda-MA-5S6	lambda-MA-3A8
V	Tg-LCb80	1st	lambda-MA-5S7	BGLB-MA-3A6
		2nd	lambda-MA-5S8	BGLB-MA-3A2

Table 2. Primer sequences for bisulfite sequencing analysis

	name	Sequences
5′ primer	LCR-MA-5S1	5′- TATAGATGTTTTAGTTTTAATAAG -3′
	ICR-MA-5S4	5′- GAATTTGGGGTATTTAAAGTTTTG -3′
	ICR-MA-5S13	5′- GGTGATTTATAGTATTGTTATTTG -3′
	lambda-MA-5S1	5′- ATTAGTAAGAAGATAGTAGTGATG -3′
	lambda-MA-5S4	5′- TTAAGTTTTGTGTGTTATTTATTA -3′
	lambda-MA-5S5	5′- GTTAAAAAGAAGAAGTAAGTATTT -3′
	lambda-MA-5S6	5′- GTGAAAGTATTGATTATTATGTTA -3′
	lambda-MA-5S7	5′- GAGGTTTATTTGTATTTATTTTTGTT -3′
	lambda-MA-5S8	5′- TATTTTTTAGTAGTATTGTAAGAGGT -3′
3′ primer	ICR-MA-3A14	5′- AAAACATAAAAACTATTATATACA -3′
	ICR-MA-3A15	5′- ACCAACCAATATAACTCACTATAA -3′
	ICR-MA-3A26	5′- CAAATTAACAAAAACATACCTAACT -3′
	BGLB-MA-3A2	5′- TTCTAACCCCACAAAAATTTATTC -3′
	BGLB-MA-3A6	5′- CCAAACCCCCTCTATTTTATATCA -3′
	lambda-MA-3A2	5′- ATACCTTATTTTTTTCTACTACAA -3′
	lambda-MA-3A3	5′- CTAAACTCCAACATATAATAACCC -3′
	lambda-MA-3A7	5′- AACCAAAATTATCTTTTTCTATCT -3′
	lambda-MA-3A8	5′- ACAACATTCTTAAATCCAATATTA -3′

RT-qPCR

Total RNA was recovered from phenylhydrazine treated anemic adult spleens (1–2 months old) of Δ5 YAC TgM using ISOGEN (Nippon Gene) and converted to cDNA using ReverTra Ace qPCR RT Master Mix with gDNA Remover (TOYOBO). Quantitative amplification of cDNA was performed with the Thermal Cycler Dice (TaKaRa Bio) using TB Green Premix EX TaqII (TaKaRa Bio). PCR primers were reported previously (27).

RESULTS

Search for proteins that bind to the 118 bp sequence within the *H19* ICR

By generating a series of TgM lines that each carry the 5′-deletion mutant of the mouse *H19* ICR sequence, we previously demonstrated that a 118 bp sequence that is located between 5′-ends of the del-6 and del-8 sequences (Figure 1A and B) is essential, both for acquiring post-fertilization methylation and for maintaining its paternally

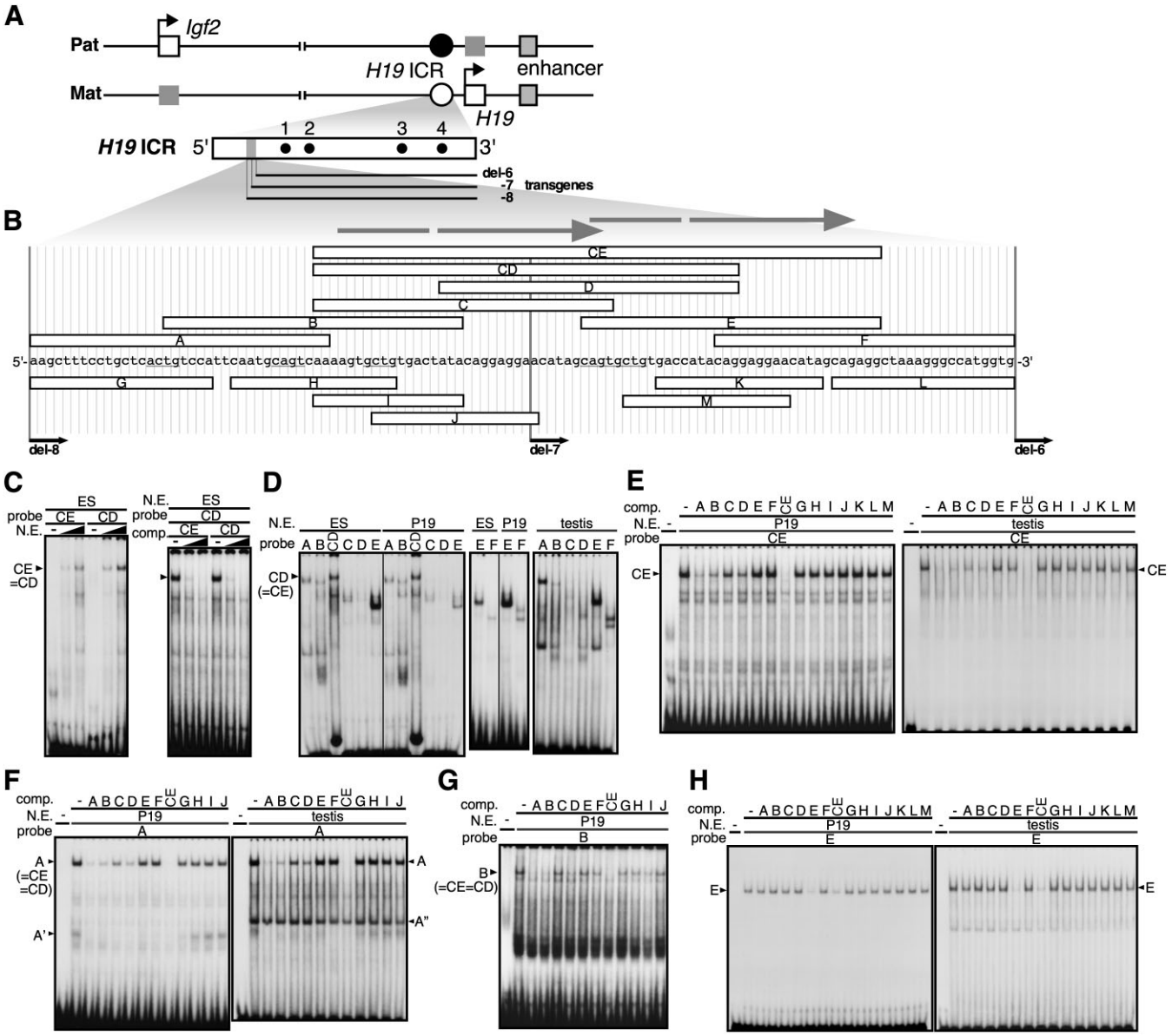


Figure 1. Search for protein binding activity to the 118 bp sequence. **(A)** Schematic representation of the mouse *Igf2/H19* locus. Monoallelic expression of paternal *Igf2* and maternal *H19* genes depends on the shared 3′ enhancer and methylation state of the *H19* ICR; the paternal allele is methylated (solid circle) while the maternal allele is unmethylated (open circle). The *H19* ICR, which contains CTCF binding sites 1–4 (dots), is located approximately 2–4 kb upstream of the *H19* transcription start site. A 118-bp sequence corresponding to the region of difference between del-6 and del-8 transgene is required for post-fertilization imprinted methylation of the *H19* ICR (28). **(B)** The 118 bp sequence and its subregions (open rectangles) used in the EMSA. Position of transgene fragment ends (thick vertical lines; del-6–8) and direct repeat sequence (thick gray arrows) is shown. **(C)** EMSA with nuclear extract (N.E.: 0, 2 and 7 μg) from mouse ES cells, and CE or CD probes (left panel). Non-labeled DNA fragments were used as competitors (20 and 100-fold molar excess) in the EMSA with 7 μg of nuclear extract (right panel). **(D)** EMSA with nuclear extracts from mouse ES, P19 embryonal carcinoma, and testis cells, and various portions of the 118 bp sequence (probes A–F and CD). **(E–H)** EMSA with nuclear extracts from P19 or testis cells, and probes CE (**E**), A (**F**), B (**G**) and E (**H**). One hundred-fold molar excess of competitors (A–M and CE) were used.

inherited hypermethylated state at the endogenous gene locus during the pre-implantation period (28). In order to identify candidate *trans*-acting factors responsible for the activity, we conducted electrophoretic mobility shift assays (EMSA). We first generated six overlapping probes (A–F) that cover the entire 118-bp sequence (Figure 1B). Because an obvious direct repeat was observed at the 5′-end of the del-7 sequence, which was partially methylated post-fertilization in TgM (28), we also generated a CE probe that covers this entire repeat sequence. Mouse embryonic stem (ES) cell nuclear extracts were incubated with radiolabeled CE probe and we detected several complexes. We named the major complex the CE complex (Figure 1C, left). This same complex also formed when the shorter CD probe was used (CD complex; Figure 1C, left), which we verified with a competition assay (Figure 1C, right), indicating that the entire direct repeat sequence is not required for the CD complex to form.

Next, A–F, along with CD probes, were individually incubated with nuclear extracts prepared from either mouse ES cells, P19 mouse embryonal carcinoma cells, or mouse testis cells (Figure 1D). Various complexes with different mobilities were observed with each probe, but cell type-specific differences were not pronounced. Additionally, neither the C, D, nor E fragment generated an apparent CD (i.e. CE) complex, at least when used as a probe, suggesting that either a long DNA motif or more than one DNA motif is being recognized for robust complex formation.

To determine if CD (CE) factor binds to other parts of the 118 bp sequence, we conducted a competition assay against the interaction between the CE probe and nuclear extracts (P19 and testis). Although less efficient than the CE fragment, the addition of either of the fragments A–D interfered with complex formation (Figure 1E), suggesting that the CD complex also forms on these sequences. To verify this result, we conducted a competition assay using fragments A (Figure 1F), B (Figure 1G), and E (negative control; Figure 1H) as probes and obtained essentially the same results as we did with the probe CE. These results suggest that the protein(s) that bound tightly to the A and CD (CE) probes and moderately to the B, C and D probes are identical, and hereafter referred to as the CD factor. Again, as the CD factor did not bind efficiently to any of the shorter fragments G–M (Figure 1E–G), its binding motif is likely to be longer or consists of multiple motifs. The other complexes, such as A', A" (in Figure 1F) and E (in Figure 1H), were not analyzed further in this study.

Identification of a protein binding motif within the 118 bp sequence

We determined the binding motif for the CD factor in fragment A by using P19 nuclear extracts. Transversion mutations (one to three nucleotides) were introduced into fragment A (m1–14, m31–33; Figure 2A) and the resulting constructs used as competitors in EMSA. As shown in Figure 2B and C, mutations in either one of two RCTG (R = A or G) motifs facing the opposite direction in fragment A significantly compromised its binding to CD (= A) factor. In other words, CD factor seems to recognize a set of RCTG motifs for strong binding, although a single motif can still be recognized, albeit with only moderate affinity (*ex.* m31 or m32 compared to m33 in Figure 2C).

To precisely determine the binding motif at single nucleotide resolution, we next used the shorter version of fragment A (m15) and its mutant derivatives (m16–40) as competitors in EMSA. We found that CD complex formation was efficiently outcompeted by including fragment A (m15) (Figure 2B–E) as this fragment harbors two RCTG motifs. Moreover, in addition to each nucleotide in the RCTG motif, the C nucleotide neighboring its 5′ end was important for recognition by CD factor (Figure 2D), which we also confirmed by using A (m40) competitor (Figure 2D) and CD probes (Figure 2E). These results suggested that a set of CRCTG is an optimal binding motif for the CD factor.

We searched for the RCTG motif within the entire 118 bp sequence and found five of them termed the motifs I–V (Figure 3A). Among these, motifs I and IV contain the optimal CRCTG sequence. The distribution of the motifs was consistent with the EMSA results shown in Figure 1E and F, in which CD factor bound to fragments A–D each contain more than one RCTG motif (Figure 3B). In contrast, fragments G–I probably did not exhibit significant binding affinity to CD factor because they each contain a single RCTG motif (H contains two but one of them is found at the end; Figure 1E and F).

To test this prediction, we replaced the C at the second nucleotide of the RCTG motif with A in each fragment [A(Δ2), B(Δ2), C(Δ2), E(Δ2) and CD(Δ3) in Figure 3B] and used them as competitors in EMSA, because such a transversion mutation most efficiently reduced their affinity for CD factor (Figure 2).

A major complex generated by probe A and nuclear extracts from P19 and ES cells in EMSA (labelled as A in Figure 3C) was outcompeted by including the A (WT), but not the mutant A(Δ2), fragment in the reaction. One of the several complexes generated with probe B (Figure 3C) exhibited similar mobility as that of the complex A and was outcompeted by the WT but not the mutant B(Δ2), fragment. Probe C generated several complexes (Figure 3D), among which only a minor band (labelled 'C') disappeared when the WT fragment, but not the C(Δ2), was used as a competitor. We therefore assumed that complexes B and C contain the CD factor, and that their weak affinity to the probes was likely due to a lack of an optimal CRCTG sequence in both fragments. In addition, motif IV in fragment C may not be fully functional because it is located at the very end of the fragment (Figure 3B; this issue will be discussed later). As we did not observe a significant difference between WT E and mutant E(Δ2) fragments in the competition reaction (Figure 3E), CD factor does not seem to bind to the fragment E. Because fragment D contains the same motifs (IV and V) as fragment E (Figure 3B), which could bind the CD factor (Figure 1E and F), the location of the motifs within the fragment may also be important for binding affinity.

Next, we performed EMSA with the CD probe and used the fragments A–C and their mutants as competitors (Figure 3F). While complex CD was disrupted by including the WT fragments (A–C and CD) as competitors in the reaction, these fragments with mutations at RCTG motifs failed to do so, indicating that CD factor binds all these fragments through the RCTG motifs.

Determination of protein binding properties to multiple RCTG motifs

The CD probe binds the CD factor most efficiently (Figure 3F) and contains three RCTG motifs (III–V; Figure 4A), among which the motif IV in the middle matches the optimal CRCTG sequence. To further clarify the mode of DNA recognition by the CD factor, we introduced distinct mutations (one to three nucleotide substitution) into the fragment CD and used them as competitors in EMSA with P19 nuclear extracts (Figure 4B). The competitive ability of each fragment summarized to the right of Figure 4A indicates that the fragment is capable of binding to the factor if at least two motifs remain intact. Additionally, motif IV appears to make the most significant contribution to the affinity. We also found that even when they contain two motifs, the competitor works less efficiently if it has insuffi-

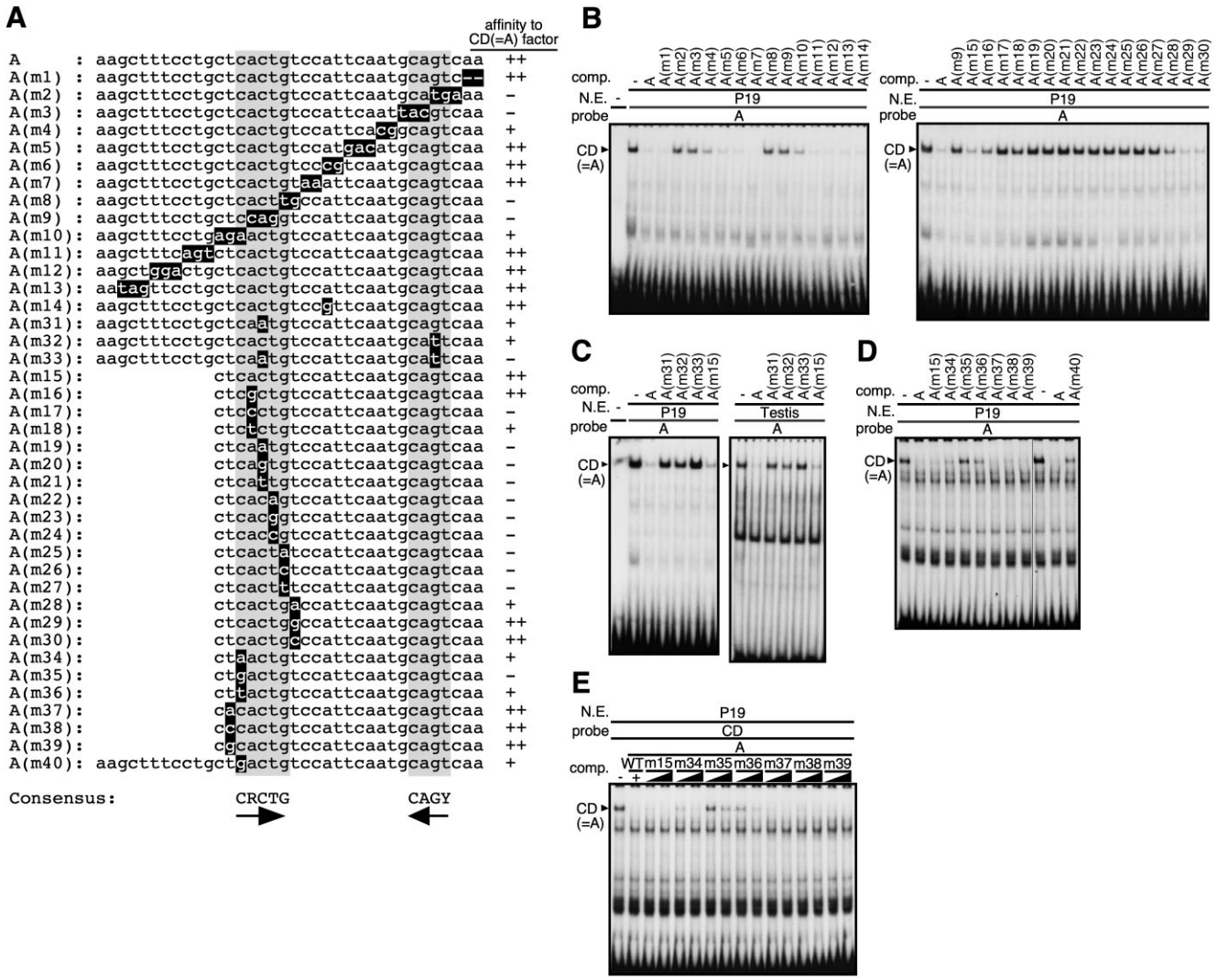


Figure 2. Search for binding motifs for CD (= A) factor in the fragment A. **(A)** Nucleotide sequence of fragment A and its mutant derivatives. Mutagenized nucleotides are inverted in black-and-white. The binding affinity of each sequence to the CD (= A) factor, which was estimated by visual-examination of the EMSA results by three individuals, is indicated by ++ (strong), + (moderate) and – (weak) to the right of each sequence. Sequences important for binding are highlighted in gray, and consensus motif is shown below the list. **(B–D)** EMSA with nuclear extracts from P19 or testis cells and probe A. One hundred-fold molar excess of competitors (mutants of A sequence) were used. **(E)** EMSA with nuclear extract from P19 cells and probe A. Fifty- and 150-fold molar excess of competitors (mutants of A sequence) were used. +: 50-fold molar excess.

ciently long flanking sequences (Figure 4A and C). Finally, the nucleotides of the CD sequence, except for the RCTG motif, were shuffled and used as competitors for EMSA (Figure 4D and E). While the scrambled oligo (scrambled-CD) bound to the CD factor with comparable affinity, the oligo with three nucleotide mutations in the RCTG motifs (scrambled-CD(Δ3)) failed to do so, indicating that RCTG is the bona fide binding motif for the CD factor.

Nuclear receptor proteins bind a direct repeat in the 118 bp sequence

We next analyzed protein factors that bind to a fragment F by EMSA (Figure 4F). When the F probe was incubated with nuclear extracts from either P19 or testis cells, two and four shifted bands were observed, respectively. Among these bands, the one with the highest mobility shift seemed to be specific binding since they were subjected to self-competition (Figure 4G). The factor was bound by fragment L, but not by fragment K, indicating that its binding site is located at 3′ portion of fragment F. In fragment L, we found a direct repeat of the aggc(t/c)a-like sequences that is a putative binding sites for nuclear receptor-type transcription factors. Among them, TR (testicular receptor) 2/4, RXR, and COUP-TF transcription factors are known to bind direct repeats of DNA Response element (AGGTCA) with one nucleotide spacing (DR1) (39–41). We therefore employed DNA fragments containing the binding sites for these factors (Epsi and RARE; (35)) and found efficient competition by EMSA (Figure 4H). We proceeded to conducted super-shift assay with antibodies against TR2 and TR4 because these DR1-binding factors are expressed in the testis (Figure 4H). Since the shifted bands either disappeared or were even further retarded, the complex appears

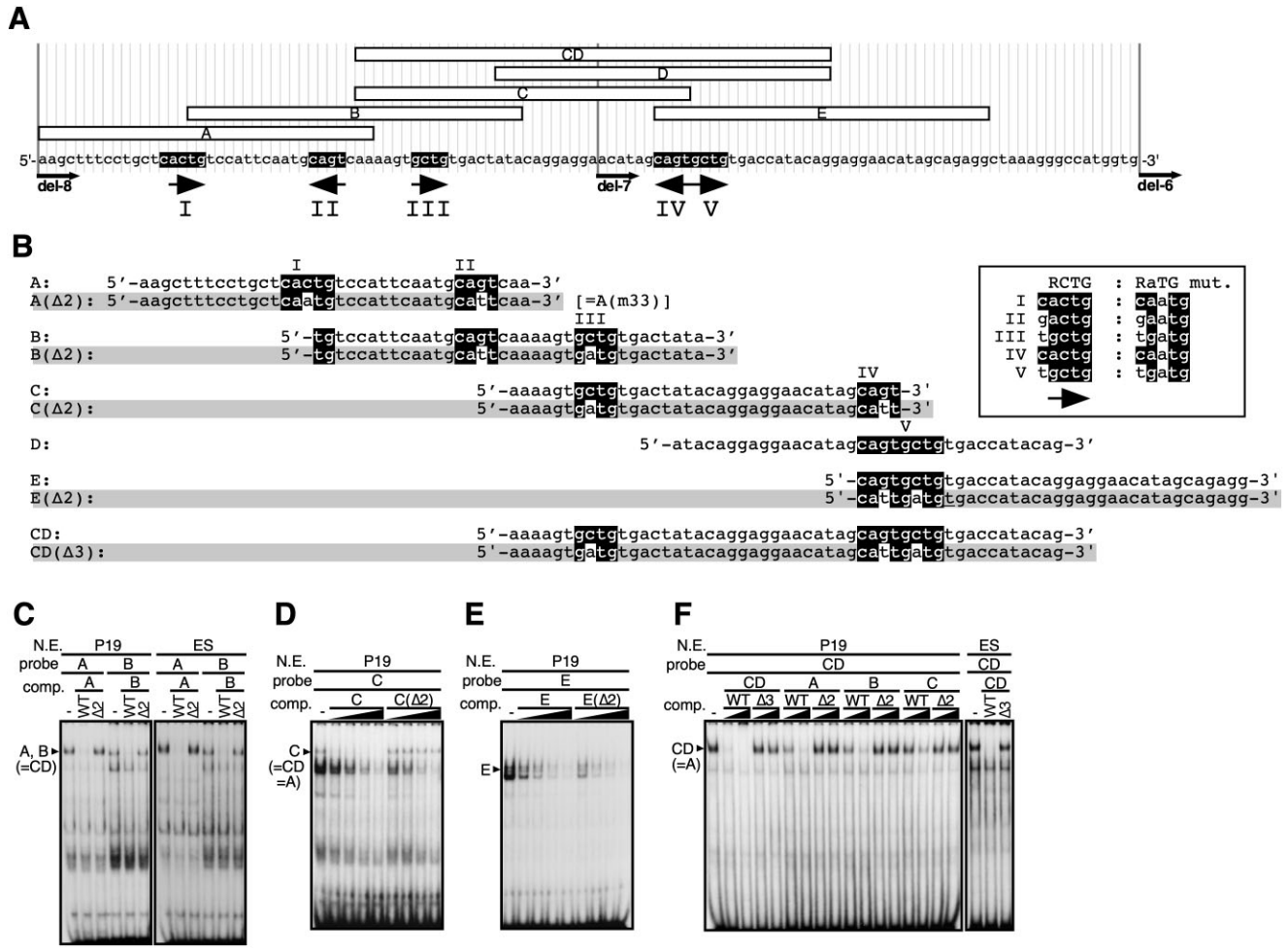


Figure 3. Search for binding motifs for CD (= A) factor in the 118 bp fragment. **(A)** The 118 bp sequence and its subregions (open rectangles) used in the EMSA. Positions of (C)RCTG motifs are shown as horizontal thick arrows (I–V). **(B)** Sequences of fragments A–E and CD, as well as their mutant derivatives [A(Δ2)–E(Δ2) and CD(Δ3)]. The (C)RCTG motifs in each fragment are inverted in black-and-white, in which mutated bases are shown in black letters on a white background. **(C–F)** EMSA with nuclear extracts from P19 or ES cells, and probes A and B (**C**), C (**D**), E (**E**) and CD (**F**). Twenty-fold (**C**) or increasing amount (100–200–500–1000-fold in (**D**), 20–50–100–200-fold in (**E**) and 20–100-fold in (**F**) molar excess of oligos were used as competitors.

to include TR2/4. Furthermore, we conducted EMSA with nuclear extracts from P19 cells in which either TR2, TR4 or both proteins were knocked down with siRNA. We found that the complex F formation efficiency was reduced in all these samples, which suggests that the TR2/4 complex binds the DR1 sequence in probe F *in vitro* (Figure 4I).

In vivo test for the function of RCTG and direct repeat sequences in YAC TgM

We subsequently determined the *in vivo* function of the five RCTG motifs, as well as of the DR1 sequence, within the 118 bp sequence of the *H19* ICR in TgM (Figure 5A). To generate a *H19* ICR Δ5 construct, we introduced a single-nucleotide mutation into each of the five RCTG motifs in the 118 bp sequence to interfere with their binding to the CD factor (Δ5; Figure 5B and C). As we have previously shown that the transgenic LCb118 sequence, within which the 118 bp sequence from the *H19* ICR is linked to the LCb (λ+CTCF+b; λ DNA harboring the CTCF and Sox-Oct motifs) sequence (28), acquired DNA methylation during post-fertilization period only when it was paternally inherited, we decided to employ this experimental setup to analyze the activity of DR1. To generate the LCb80 sequence, a 38 bp sequence, containing the DR1, was deleted from the 118 bp sequence. The resulting 80 bp sequence (underlined in Figure 5B) was linked to the LCb sequence (LCb80; Figure 5C and (33)). To employ transgene coplacement strategy, these two fragments (Δ5 and LCb80) were independently floxed by a set of hetero-specific *loxP* sequence, linked in tandem and introduced into human β-globin yeast artificial chromosome (Δ5-LCb80 YAC; Figure 5C).

Two TgM lines, each carrying a single-copy YAC, were generated by injecting the YAC DNA into fertilized mouse embryos. Southern blot analysis of high molecular weight thymic DNA of the TgM revealed that one of the two lines (line 723) lacks the adjacent region 3′ to the δ-globin gene, while the other (line 753) carries an intact copy (Figure 5C and D). These mice were crossed with Cre recombinase TgM to induce Cre-loxP recombination in oocytes, resulting in TgM sub-lines carrying either the Δ5 or LCb80 sequences (Figure 5E). The occurrence of the desired recom-

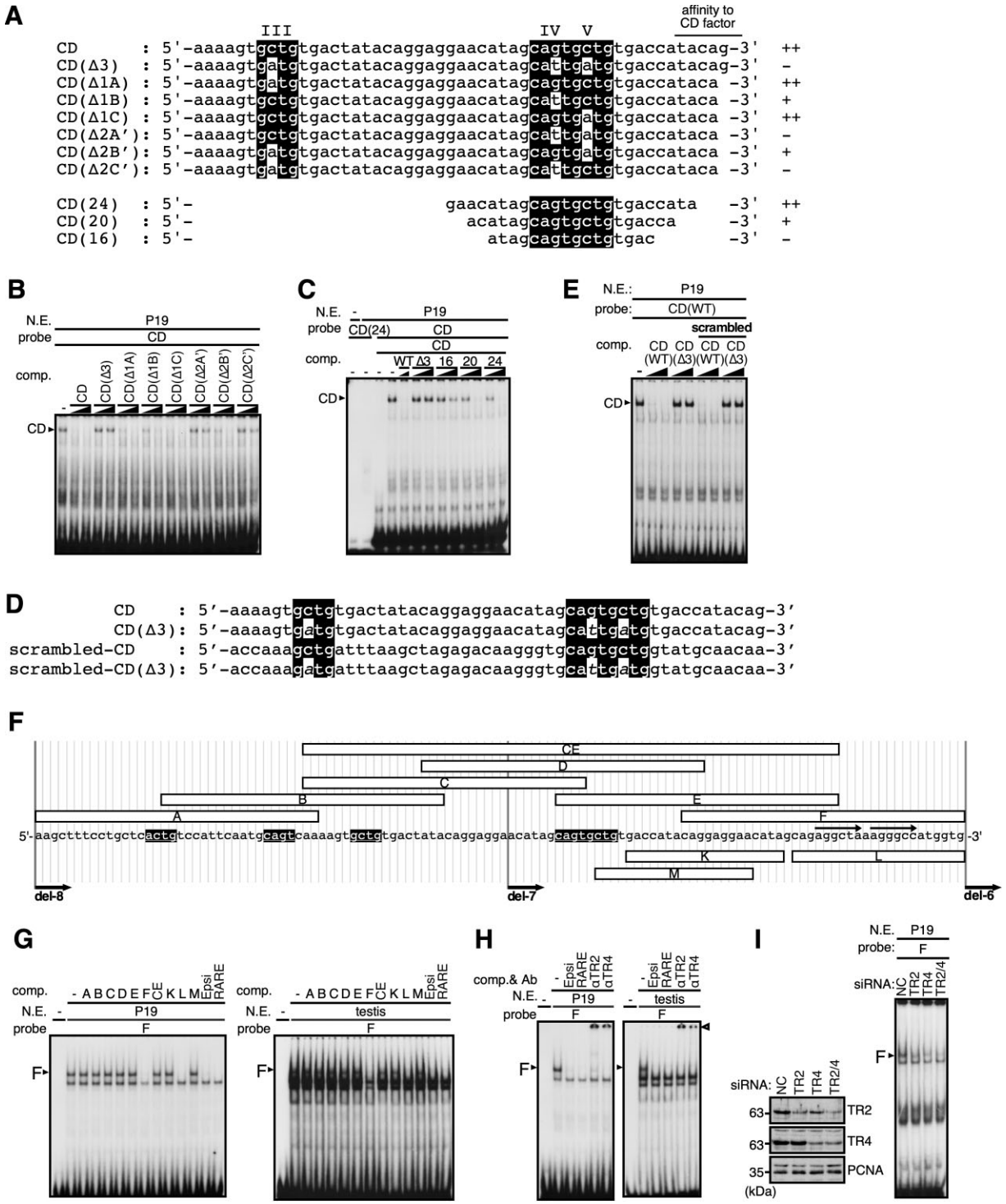


Figure 4. Search for binding motifs for CD (= A) factor in the fragment CD. **(A)** Sequences of fragment CD and its mutant derivatives. The (C)RCTG motifs in each fragment are inverted in black-and-white, in which the mutated bases are reverted. The binding affinity of each sequence to the CD (= A) factor, which was estimated by visual-examination of the EMSA results by three individuals, is indicated by ++ (strong), + (moderate) and – (weak) to the right of each sequence. **(B, C)** EMSA with nuclear extract from P19 cells and probe CD. Forty- and 200-fold molar excess of oligos were used as competitors. **(D)** Sequences of fragment CD and its scrambled mutants. Sequence composition (*i.e.* the number of each nucleotides) is maintained even after shuffling the sequence. **(E)** EMSA with P19 cell nuclear extract and probe CD. Twenty- and eighty-fold molar excess of oligos were used as competitors. **(F)** Putative binding site for nuclear receptor-type transcription factors (*i.e.* direct repeat sequence) is indicated by a pair of arrows. **(G–I)** EMSA with nuclear extracts from P19 or testis cells and probe F. One hundred-fold molar excess of oligos were used as competitors. Nucleotide sequences of 'Epsi' and 'RARE' oligos each carrying binding motifs for TR2/TR4 are as follows: Epsi, 5'-CTGAGGACACAGGTCAGCCTTGACCAATGACTTTTA-3' and RARE, 5'-TTGCTGTGACCTCTGCCCTTCTAGCCTCT-3' (only the sequence of one strand is shown and binding motifs are underlined). In Figure 4H, the super-shift bands observed in the presence of antibodies against either TR2 or TR4 are indicated by an open triangle. In Figure 4I, nuclear extracts were prepared from P19 cells transfected with siRNA duplexes against TR2, TR4, or both. The decrease in protein levels in these cells was verified by western blotting. NC; non-targeting control siRNA.

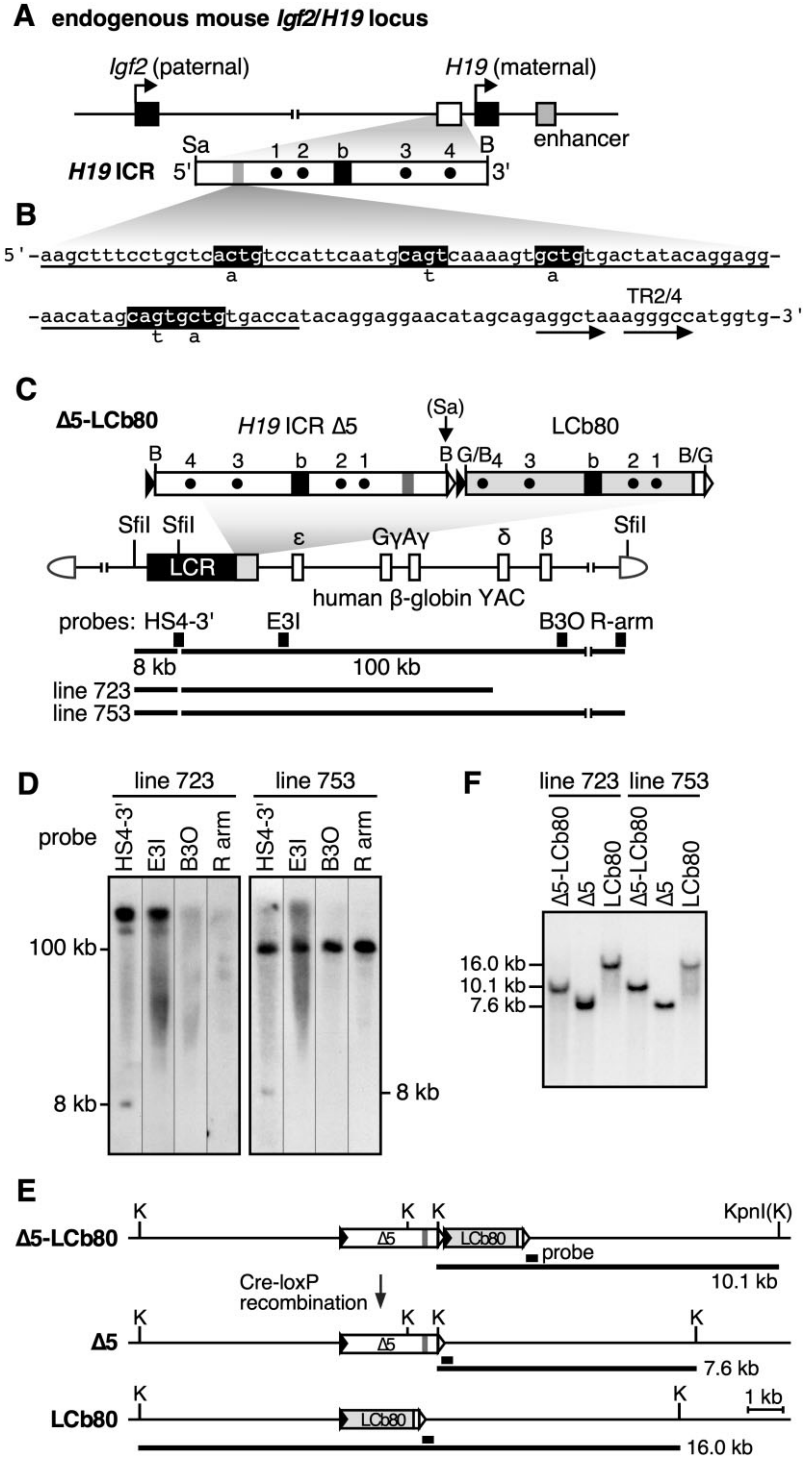


Figure 5. Generation of YAC TgM. **(A)** Structure of the mouse *Igf2/H19* locus. The *H19* ICR is contained within a 2.9-kb *Sac*I (Sa)-*Bam*HI (B) fragment, in which CTCF binding sites and the 'b' region (33) are indicated by dots (1-4) and a filled box, respectively. **(B)** The 118 bp sequence in the *H19* ICR. (C)RCTG motifs are highlighted with their mutated sequences in the Δ5 mutant underneath. The sequences included in the LCb80 are underlined and TR2/4 binding site is shown by tandem arrows. **(C)** Structure of the 150-kb human β-globin locus YAC. The LCR and β-like globin genes are denoted as black and open boxes, respectively. Each of the *H19* ICR Δ5 (open rectangle; 2.9 kb) and LCb80 (gray; 2.4 kb) fragment was floxed by a pair of loxP sites [loxP5171 (solid triangles) and loxP2272 (open)], tandemly arranged and introduced 3' to the LCR for employing co-placement strategy. The expected *Sfi*I restriction enzyme fragments (thick lines) generated from the YAC transgene and probes (filled rectangles) are shown beneath the map. **(D)** Long range structural analysis of the Δ5-LCb80 YAC transgene. DNA from thymus cells was digested with *Sfi*I in agarose plugs and separated by pulsed-field gel electrophoresis, and Southern blots were hybridized separately to probes. **(E,F)***In vivo* Cre-loxP recombination to derive Δ5 or LCb80 TgM. Recombination between two loxP5171 sites (solid) in the parental Δ5-LCb80 transgene, for example, would generate LCb80 allele, during which one of the loxP2272 sites (open) is concomitantly removed to prevent further recombination. Tail DNA from parental and daughter YAC-TgM sublines was digested with *Kpn*I and analyzed by Southern blotting using the probe shown in (**E**).

bination event was verified by Southern blot analysis of tail somatic DNA from sub-line animals (Figure 5E and F).

Five-nucleotide mutation in RCTG motifs abolishes imprinted DNA methylation of the *H19* ICR

We first analyzed the methylation state of the transgenic *H19* ICR fragment with a Δ5 mutation (Tg-Δ5) by Southern blotting (Figure 6A). While the maternally inherited mutant *H19* ICR sequence was correctly hypomethylated in tail somatic cells as expected, the same sequence became hypomethylated even after paternal transmission in most cases (14/14 and 16/18 in lines 723 and 753, respectively; Figure 6B-E). Similar to the case of WT *H19* ICR in TgM, the Δ5 mutant sequence was devoid of methylation in the testis of TgM (Figure 6F and G). Therefore, the Δ5 mutation is likely to hinder post-fertilization methylation of the *H19* ICR.

Next, we pooled tail somatic cell DNA from animals carrying hypomethylated (as judged by Southern blot analysis) paternal and maternal Δ5-Tg, separately. The pooled DNA was analyzed by bisulfite sequencing (Figure 6H). Consistent with our Southern blotting results, paternally inherited Tg DNA in tail somatic cells were hypomethylated (top) and indistinguishable from the maternally inherited transgene in terms of methylation pattern (middle). The Tg sequence close to the Δ5 mutation was moderately methylated after either paternal or maternal transmission, which we also observed for the WT *H19* ICR Tg (Figure 6I) because this region lies outside the DMR. In addition, because the sequence was not methylated in either sperm, one-cell embryos, or two-cell stage embryos (Figure 6H), DNA methylation at this region was likely acquired after the implantation stage by non-allele specific *de novo* DNA methylation activity.

Expression of human β-globin genes is activated by the LCR enhancer and is not normally subject to genomic imprinting. We have previously demonstrated that introducing the wild-type mouse *H19* ICR fragment between the LCR and the β-globin genes (in the same position as Δ5 in this study) of a transgenic human β-globin locus YAC recapitulated its imprinted expression in DNA methylation-dependent fashion. We found that the paternally inherited transgenic β-globin gene was highly expressed, whereas the maternally inherited copy was repressed as CTCF proteins bound to the hypomethylated maternal *H19* ICR insulate the gene from the LCR's enhancer activity (Figure 6J, WT, (24)). As decreased methylation of the paternal *H19* ICR in Δ5 TgM was expected to allow its binding to CTCF and insulator formation, we analyzed the transgene's expression state in Δ5 TgM (Figure 6J, Δ5), and observed low expression of the β-globin gene regardless of its parental origin, demonstrating that the five nucleotides are essential for genomic imprinting of the transgene.

A direct repeat may regulate imprinted DNA methylation of the *H19* ICR in concert with RCTG motifs

We went on to determine the methylation state of the transgenic LCb80 fragment by Southern blotting (Figure 7A). As shown in Figure 7B–E, the maternally inherited LCb80 transgene in tail somatic cells of TgM was hypomethylated, as we anticipated. On the other hand, paternally inherited Tg-LCb80 sequence in some of the animals was either partially or fully hypomethylated (4 and 4, respectively, out of 16 in line 723; and 1 and 9, respectively, out of 19 in line 753). We have previously shown that the paternally inherited transgenic LCb118 sequence is DNA-methylated during the post-fertilization period in almost all the cases (28). In addition, the Tg-LCb80 sequence was hypomethylated in the testis (Figure 7F and G), as was the LCb118 sequence in TgM. These findings indicate that post-fertilization DNA methylation of LCb118 was affected by removing the 38 bp sequence. They also suggest that the 38 bp sequence may function in concert with the rest of the 118 bp sequence.

To test the above predictions, we deleted the DR1-containing sequence from the WT Tg *H19* ICR in YAC TgM by genome editing (*H19* ICR Tg-Δ36; Figure 8A), and generated two YAC TgM lines (lines 53 and 55) that harbor the same 36 bp deletion (dotted lines in Figure 8B). Due to limitations caused by the PAM sequence availability, we could only design a set of guide RNAs (underlined in Figure 8B) that would eliminate a 36-bp, instead of 38-bp (= 118 minus 80), sequence. We analyzed the methylation state of the mutant Tg sequence by Southern blotting following either paternal or maternal inheritance (Figure 8C and D). Although some degree of heterogeneity was noted between the two lines, both showed loss of methylation in the paternally inherited mutant Tg (severe and moderate in lines 53 and 55, respectively; Figure 8C and D). Hypomethylated paternal (pat., top panel in Figure 8E) and hypermethylated paternal (pat., middle in Figure 8E) samples, whose methylation state was determined by Southern blotting, were further analyzed by bisulfite sequencing to verify their methylation state. Both Southern blotting and bisulfite sequencing revealed that the maternally inherited mutant *H19* ICR fragment was hypomethylated (Figure 8C–E).

The RCTG motif is present in the IG-DMR of the mouse *Dlk1-Dio3* gene locus

We recently reported that the rat *H19* ICR, orthologous to the mouse *H19* ICR, did not acquire post-fertilization methylation in its TgM following paternal transmission (42). As shown in Figure 9A, the number and location of RCTG motifs within the 118 bp sequence (113 bp in rat) are not highly conserved between mouse and rat. In addition, no distinct direct repeat motif was found in the rat sequence. To test if the rat 113 bp sequence binds CD and TR2/4 *in vitro*, we performed EMSA with a rat-AC probe that contains the three RCTG motifs, and rat-DR probe, corresponding to the mouse L probe covering the TR2/4 binding motif (Figure 9A). The rat-AC sequence competed less efficiently than the CD(WT) sequence for binding to CD factor in P19 nuclear extract, suggesting that the rat sequence has moderate affinity for the CD factor (Figure 9B). On the other hand, the rat-DR, as well as the rat-AC sequences, did not compete with the mouse F probe to bind to TR2/4, which suggests that the rat DR sequence cannot bind to TR2/4 (Figure 9C). Post-fertilization imprinted methylation may not have been reproduced in the transgenic

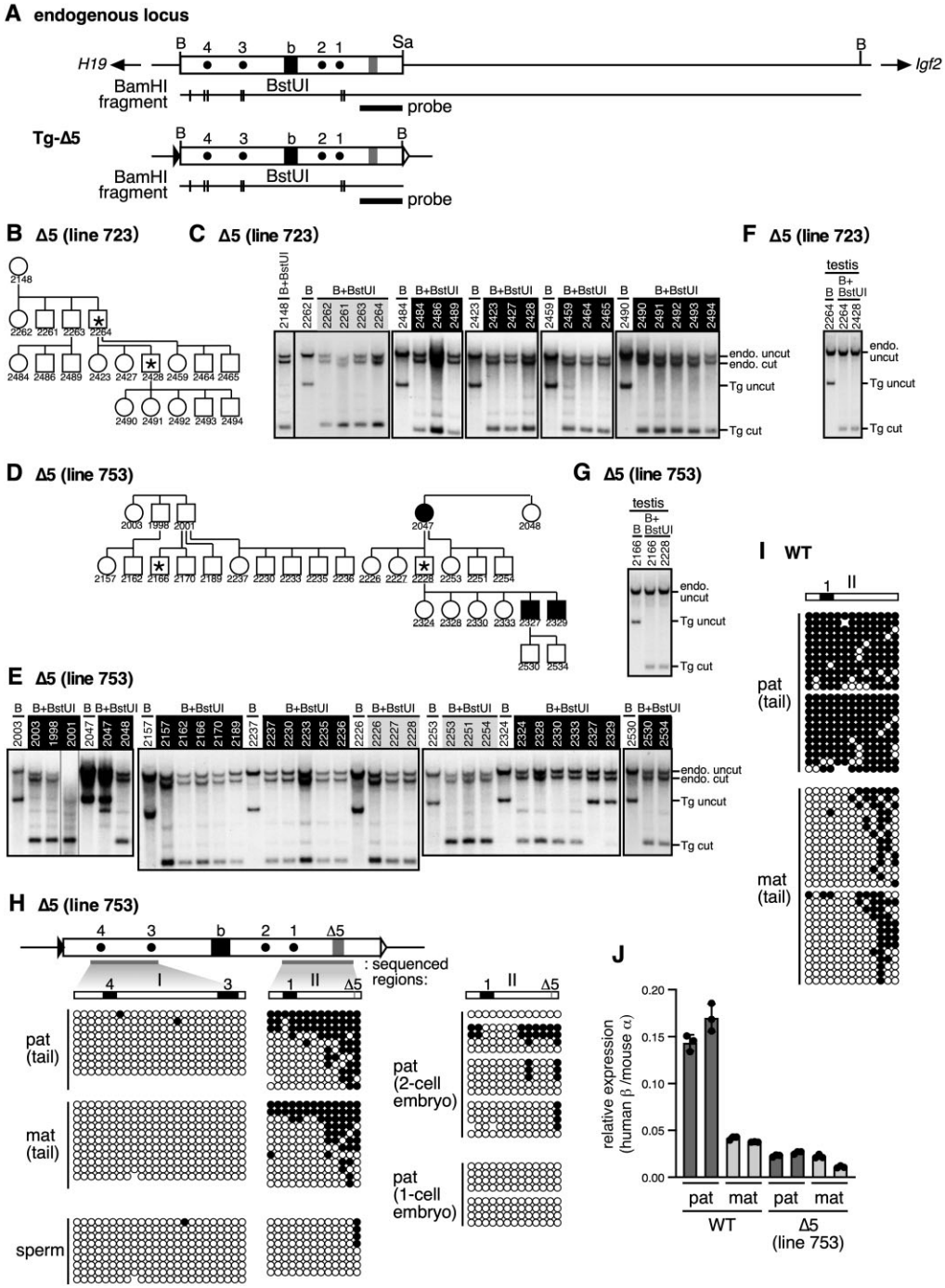


Figure 6. DNA methylation state of the Δ5 *H19* ICR transgene. **(A)** Partial restriction enzyme maps of the endogenous *H19* locus and the β-globin YAC transgene with the inserted Δ5 *H19* ICR fragment. Methylation-sensitive *Bst*UI sites in the *Bam*HI fragments are displayed as vertical lines beneath each map. The probe used for Southern blot analysis in (**C**, **E**–**G**) is shown as a filled rectangle. B; *Bam*HI, Sa; *Sac*I. **(B, D)** Pedigree of Δ5 *H19* ICR TgM lines 723 and 753 those were analyzed in (**C**, **F**) and (**E**, **G**, **H**), respectively. Male and female individuals are represented as rectangles and circles, respectively. Filled or open symbols indicate hyper- or hypo-methylated state of Δ5 fragment in each TgM, which was determined by visual examination of the Southern blot results in (**C**, **E**) by three individuals. Testis samples in (**F**, **G**) were obtained from male individuals marked by stars. Tail DNA from underlined animals was pooled according to the transgene's parental origin and analyzed by bisulfite sequencing in (**H**). **(C, E–G)** Southern blot results showing DNA methylation state of the Δ5 *H19* ICR fragment in tail somatic cells (**C**, **E**) and testis (**F**, **G**) of YAC-TgM. ID numbers of individuals inheriting the transgene maternally and paternally are highlighted in gray and black colors, respectively, in (**C**, **E**). endo.; endogenous locus, Tg; transgene. **(H)** DNA methylation state of the Δ5 *H19* ICR fragments in tail somatic cells, sperm, 2-cell embryos, and 1-cell embryos of YAC-TgM that inherited the transgenes either paternally (pat) or maternally (mat), was analyzed by bisulfite sequencing. Regions analyzed are indicated by gray bars beneath map. **(I)** DNA methylation state of the WT *H19* ICR fragments (region II) in tail somatic cells of YAC-TgM that inherited the transgenes either paternally (pat) or maternally (mat) was determined by bisulfite sequencing. **(J)** The relative expression levels of the human β-globin gene, after normalization to that of the endogenous mouse α-globin (reference) gene were determined by RT-qPCR. The average and standard deviation (S. D.), determined by three reactions, are depicted for each category of animals (Tg and its parental origin).

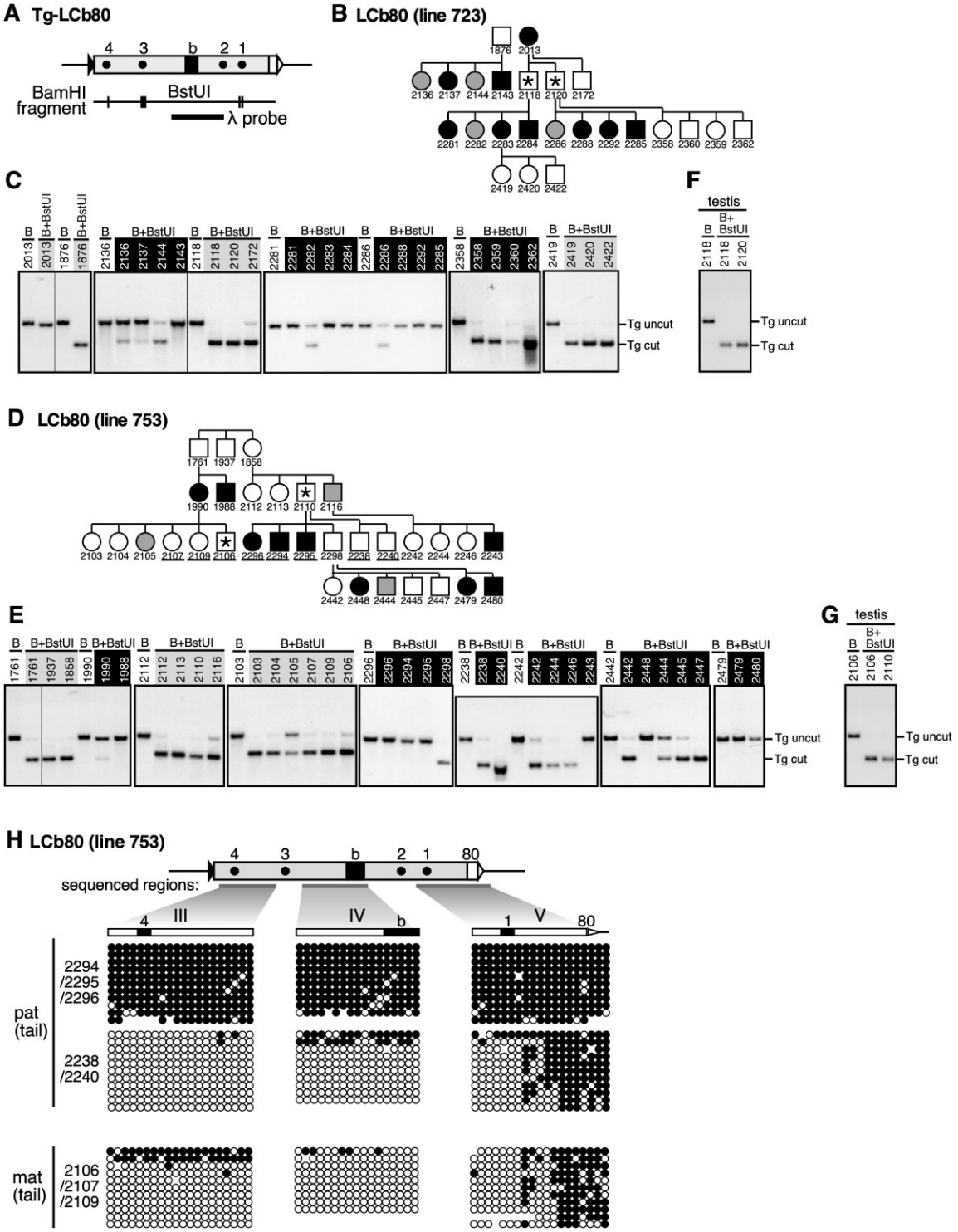


Figure 7. DNA methylation state of the LCb80 transgene. **(A)** A partial restriction enzyme map of the LCb80 fragment in the β-globin YAC transgene. Methylation-sensitive *Bst*UI sites in the *Bam*HI fragment are displayed as vertical lines beneath the map. The probe used for Southern blot analysis in (**C**, **E**–**G**) is shown as a filled rectangle. **(B,D)** Pedigree of LCb80 TgM lines 723 and 753 those were analyzed in (**C**, **F**) and (**E**, **G**, **H**), respectively. Male and female individuals are represented as rectangles and circles, respectively. Filled, gray, or open symbols indicate hyper-, partially-, or hypo-methylated state of LCb80 fragment in each TgM, which was determined by visual examination of the Southern blot results in (**C**, **E**) by three individuals. Testis samples in (**F**, **G**) were obtained from male individuals marked by stars. Tail DNA from underlined individuals were pooled according to the transgene's parental origin and their methylation levels estimated from Southern blot results, and was analyzed by bisulfite sequencing in (**H**). **(C, E–G)** Southern blot results showing DNA methylation state of the LCb80 fragment in tail somatic cells (**C**, **E**) and testis (**F**, **G**) of YAC-TgM. ID numbers of individuals inheriting the transgene maternally and paternally are highlighted in gray and black, respectively, in (**C**, **E**). Tg; transgene. (**H**) DNA methylation state of the LCb80 fragment in tail somatic cells of TgM that inherited the transgene either paternally (pat) or maternally (mat) was analyzed by bisulfite sequencing. Regions analyzed are indicated by gray bars beneath the map.

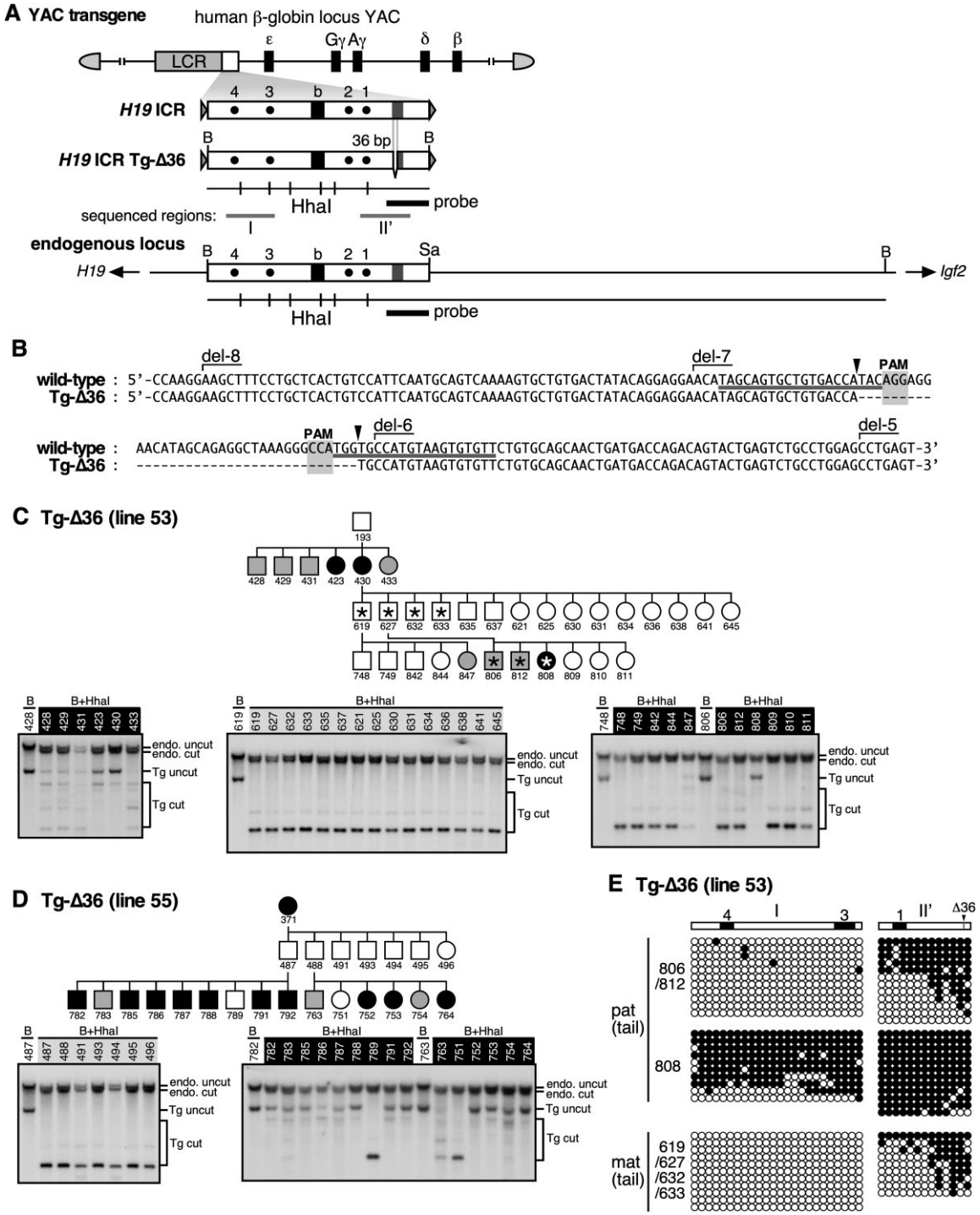


Figure 8. DNA methylation state of the Δ36 *H19* ICR transgene. **(A)** Structure and partial restriction enzyme maps of the β-globin YAC transgene carrying the WT or mutant *H19* ICR fragments (2.9 kb), in which the 36-bp sequence was internally deleted by CRISPR/Cas9 genome editing (*H19* ICR Tg-Δ36), and the endogenous *H19* locus. Methylation-sensitive *Hha*I sites in the *Bam*HI fragments are displayed as vertical lines beneath each map. The probe used for Southern blot analysis in (**C**, **D**) is shown as a filled rectangle, and regions analyzed by bisulfite sequencing in (**E**) are indicated by gray bars. B; *Bam*HI, Sa; *Sac*I. **(B)** Sequence alignment of wild-type and the mutant (Δ36) *H19* ICRs. Protospacer-adjacent motif (PAM) and gRNA sequences are shaded and underlined, respectively. Cleavage sites predicted by PAM locations (arrowheads), as well as the end positions of del-6–8 fragments (28) are shown. **(C, D)** Pedigree and DNA methylation state of the Δ36 fragment in TgM lines 53 (**C**) and 55 (**D**), respectively. In pedigrees, male and female individuals are represented as rectangles and circles, respectively. Filled, gray, or open symbols indicate hyper-, partially- or hypo-methylated state of Δ36 fragment in each TgM, which was determined by visual examination of the Southern blot results by three individuals. Tail DNA from individuals marked by stars was pooled according to the transgene's parental origin and their methylation levels estimated from Southern blot results, and was analyzed by bisulfite sequencing in (**E**). ID numbers of individuals inheriting the transgene maternally and paternally are highlighted in gray and black colors, respectively. endo.; endogenous locus, Tg; transgene. **(E)** DNA methylation state of the Δ36 fragments in tail somatic cells that inherited the transgene either paternally (pat) or maternally (mat).

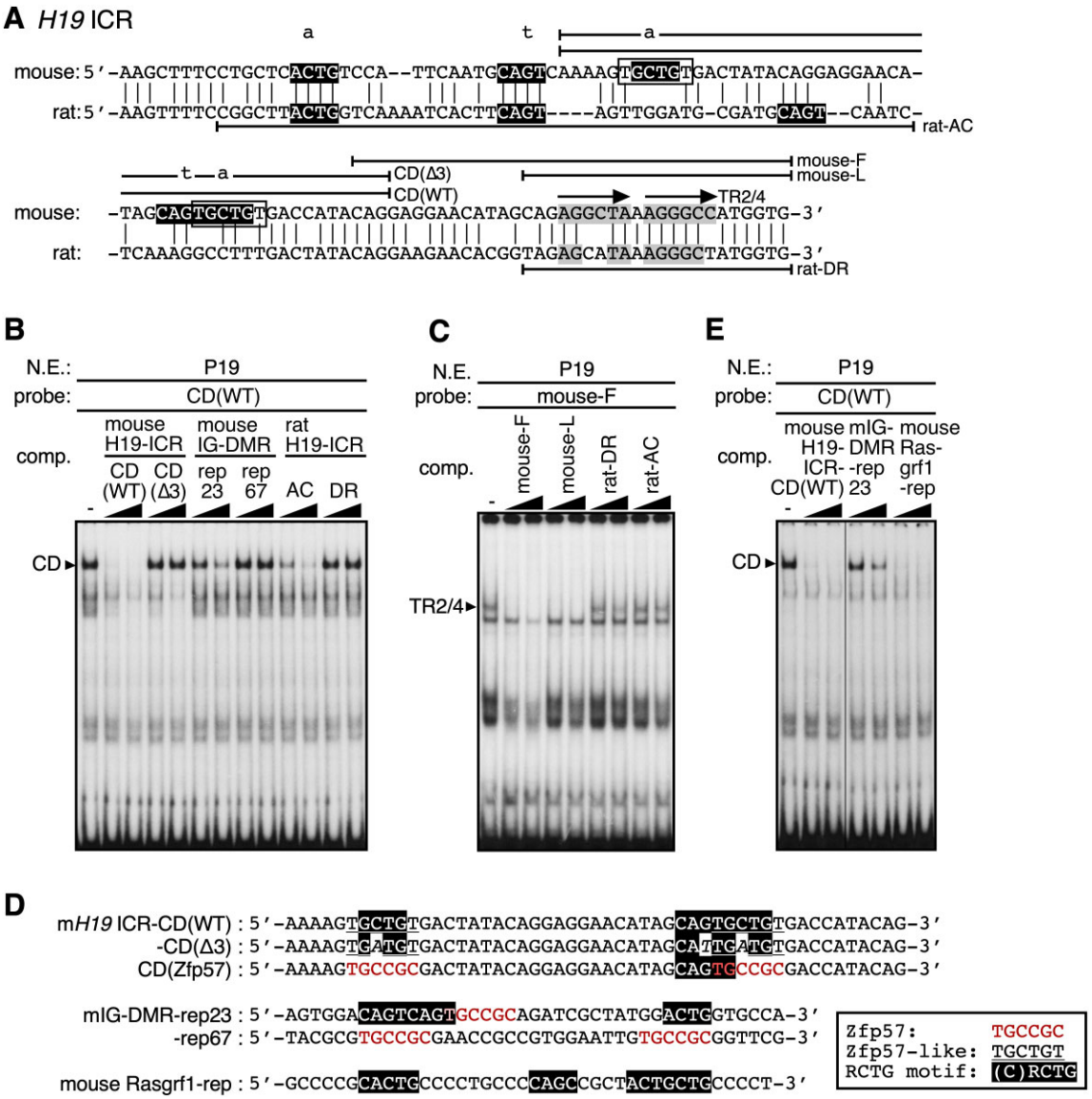


Figure 9. Ability of the CD factor to bind various DNA sequences. **(A)** Comparison between mouse 118 bp and rat 113 bp sequences. Identical nucleotides are denoted by vertical lines. The (C)RCTG motifs are inverted in black-and-white, while Zfp57-like motifs are shown in rectangles. Sequence portions used as probes in EMSA are indicated by horizontal lines. The Δ5 mutations are shown in lowercase letters above the mouse sequence. **(B, E)** EMSA with P19 cell nuclear extract and probe CD(WT). Twenty- and eighty-fold molar excess of oligos were used as competitors. **(C)** EMSA with P19 cell nuclear extract and probe F. Ten- and 40-fold molar excess of oligos were used as competitors. **(D)** RCTG-motif-containing sequences from the mouse *H19* ICR, the repeat within the IG-DMR (*Dlk1–Dio3* gene locus), and the Sp4 repeat (*Rasgrf1* gene locus). The (C)RCTG consensus motifs are inverted in black and white, while Zfp57-binding consensus motifs are shown in red. Zfp57-binding consensus-like motifs (underlined) in the CD(WT) sequence were mutated to generate Zfp57-binding consensus motifs within the CD(Zfp57) sequence.

rat *H19* ICR due to moderate binding affinity to the CD factor, an absence of TR2/4 binding motifs, or both.

The IG (intergenic) -DMR sequence (Supple. Figure 1A) is one of three paternally methylated DMRs. In the above report, we also found that the mouse IG-DMR fragment exhibits post-fertilization imprinting, as was observed in the mouse *H19* ICR in TgM (42). Since removing the repeat sequence, comprising seven repeats of a 24 bp unit, from the endogenous mouse IG-DMR results in loss of DNA methylation on the paternally inherited allele without changing its methylation in sperm (43), the repeat sequence may be responsible for acquisition of post-fertilization imprinting.

As the repeat sequence contains four consensus Zfp57 binding and five (C)RCTG motifs (Supple. Figure 1A, bottom), we conducted EMSA with probes containing either Zfp57 plus RCTG motifs (mIG-DMR-rep23: Figure 9D; R2 + R3 sequences in Supple. Figure 1A) or only Zfp57 motifs (rep67: Figure 9D; R6 + R7 sequences in Supple. Figure 1A). Recombinant GST-Zfp57 fusion protein expressed in bacteria bound to the methylated rep67 probe, as expected (Supple. Figure 1B). We found that Zfp57 protein present in the P19 nuclear extract also bound the methylated but not the unmethylated rep67 probes (Supplementary Figure 1D). We verified the binding activity of the Zfp57 from

P19 nuclear extract using a competitive binding experiment with CD(Zfp57) sequence in which Zfp57-binding consensus motifs were generated by mutating Zfp57-binding consensus-like motifs in CD(WT) (Figure 9D and Supple. Figure 1C; (27)). As shown in Figure 9B and E, CD factor bound to the *H19* ICR-CD(WT) probe was outcompeted by the mIG-DMR-rep23 sequence but not by the rep67 sequence, suggesting that the RCTG motifs found within the repeat sequence of the mIG-DMR may help confer post-fertilization imprinted methylation.

DISCUSSION

In this study, we investigated candidate 118-bp binding factors and identified one activity that recognize RCTG motifs, five of which are present in the 118-bp region. A five-nucleotide mutation (Δ5) that disrupts the factor's binding to the 118-bp sequence *in vitro* caused loss-of-methylation of the paternally inherited *H19* ICR *in vivo*, i.e. in TgM (Figure 6). Consequently, the expression of the paternally inherited β-globin transgene *in cis* was suppressed, just as with the maternally inherited allele (Figure 6J). Therefore, the binding activity identified in this study seems to be involved in the regulation of *H19* ICR methylation and genomic imprinting via the 118 bp sequence; however, some of the five mutated nucleotides may form part of the distinct recognition motif for other unidentified factors, in which case inhibition of this factor's binding to the 118-bp sequence causes loss of genomic imprinting.

The CD factor was more tightly bound when multiple RCTG motifs were present (Figure 2 and Figure 4A–C) as well as when a C is located immediately upstream of the RCTG motif (Figure 2D, E). Of the five RCTG motifs in the 118 bp sequence, two (one of them is the CRCTG) are located within the del-7 transgene that exhibited partial methylation activity upon paternal transmission in our earlier study (28). Therefore, the difference in imprinted methylation activity, *i.e.* partial in the del-7 and robust in the del-8 transgenes, may be explained by the number of (C)RCTG motifs they carry.

We previously found that the IG-DMR sequence of the mouse *Dlk1–Dio3* locus can acquire paternal-allele-specific methylation in TgM after fertilization (42). Because a number of consensus Zfp57-binding motifs are present in the repeat sequence (Supplementary Figure 1A) and are bound by Zfp57/445 *in vivo* (32), these factors may be responsible for the observed activity. However, while the involvement of these factors in maintaining post-implantation methylation has been established, their function during the pre-implantation period is not fully understood. Additionally, we detected CD factor binding to the repeat sequence (rep23) by EMSA (Figure 9B, E) in this study. Therefore, it is possible that five (C)RCTG motifs in the repeat sequence (Supplementary Figure 1A) and the CD factor may control post-fertilization *de novo* methylation also at the IG-DMR via the same mechanism in the *H19* ICR.

The gametic acquisition of DNA methylation within the *Rasgrf1* DMR, another paternally methylated DMR, is thought to involve a piRNA pathway entailing transcription of a long terminal repeat (LTR)-type retrotransposon sequence found within the DMR (44). Deletion of a nearby repeat sequence (45), which is a predicted transcription regulatory element, results in loss of DNA methylation in sperm (46). Moreover, when the repeat sequence was conditionally removed soon after fertilization, the methylation level of the paternal allele, previously established in the sperm, declined (47). Therefore, the repeated sequence may also be required to maintain DNA methylation of the paternal allele during the post-fertilization period via its *de novo* methylation activity. In accordance with this notion, several (C)RCTG motifs are found within and around the repeat sequence (Supple Figure 1E). When tested in EMSA, they work as strong competitors, as did the CD(WT) sequences (Figure 9E).

Thus, the mechanism underlying methylation induction through CD factor binding to (C)RCTG motifs in pre-implantation embryos may be conserved among three paternally methylated imprinted mouse loci. Nevertheless, although the *H19* ICR lacking the 118-bp sequence (116-bp to be exact) lost post-fertilization imprinted methylation activity *in vivo* (28), there are still many RCTG (62 sites) and CRCTG motifs (15 sites) in the remaining part of the sequence. Therefore, it is unlikely that imprinted methylation of the *H19* ICR is regulated solely by the presence of these sequences (this issue is further discussed below).

Disrupting the five RCTG motifs in the transgenic *H19* ICR (Δ5) resulted in its near-complete loss of methylation in the paternal allele (Figure 6), indicating that these motifs are essential for post-fertilization methylation imprinting. However, unlike the LCb118, which is accompanied by full imprinted methylation activity (28), the LCb80, in which an 80 bp sequence containing all the five RCTGs was linked to an LCb fragment, was only partially imprinted (Figure 7). Therefore, the 80 bp sequence (i.e. the five RCTG motifs) alone is insufficient for efficient DNA methylation, and may function in concert with the remaining 38 bp sequence. A short direct repeat motif bound by the TR2 (Nr2c1) and TR4 (Nr2c2) (48) was found within a 38 bp sequence (Figure 4F, H and I). Since TR2/TR4 has been reported to interact with Trim28 (49), which is involved in keeping the *H19* ICR methylated (31), the TR2/TR4 may recruit Trim28 and DNA methyltransferases in the pre-implantation embryo in concert with the CD factor. Our finding in the Δ36 TgM that methylation level of the *H19* ICR transgene with a 36 bp deletion was partially reduced after paternal inheritance is also consistent with our hypothesis that action of the direct repeat sequence and the five RCTG motifs are cooperative (Figure 8).

As noted earlier, even though numerous RCTG motifs are present in other regions of the *H19* ICR, mutation of only five motifs within the 118 bp sequence completely disrupted post-fertilization imprinted methylation. Thus, either the frequency of the RCTG motifs, their position and polarity, their relative arrangement with other elements such as direct repeat, or all of the above factors may be important for their stable binding to regulatory factors required for methylation. The lack of post-fertilization methylation imprinting in the TgM of the rat *H19* ICR sequence is consistent with this notion, as only three RCTG and no TR2/TR4 motifs are found within the rat 113 bp region. Whether these motifs are commonly conserved and functional in the ICRs of other imprinted loci, such as the mouse *Dlk1-Dio3* and *Rasgrf1* loci, as well as the human *H19* ICR sequence (8.8 kb, (42)), remains to be determined.

The results of our combination of *in vitro* and *in vivo* experiments strongly suggest that allele-specific DNA methylation of the *H19* ICR after fertilization requires specific DNA motifs and protein factors that recognize them. One possible function of these factors in the post-fertilization imprinted methylation mechanism is to recruit DNA methyltransferases specifically to the *H19* ICR in early embryos immediately after fertilization. Since post-fertilization DNA methylation of the *H19* ICR occurs even without DNA methylation in the sperm, these factors may also be responsible for depositing epigenetic signatures in the germline used to distinguish parental alleles after fertilization. It was recently reported that in germ cells and pre-implantation embryos, H3K9me3 modification is enriched in DNA-methylated ICRs, including the *H19* ICR. Moreover, experimental depletion of this histone modification in early embryos reportedly reduces DNA methylation levels in the ICRs (50). If this histone modification (or other unidentified modifications) is used to distinguish the parental alleles and guide imprinted DNA methylation after fertilization, the CD factor may recruit histone-modifying enzymes to the *H19* ICR in the germline via a 118 bp sequence. Future work is needed to determine the identity of the CD factor present in P19, testis and ES cell extracts, and if they, CD and TR2/4 factors, are the bona fide regulator in mouse. This will provide clues to the mechanism underlying maintenance of ICR methylation in the pre-implantation embryo, which is crucial to the execution of genomic imprinting.

DATA AVAILABILITY

Raw data, such as gel images, are available on request.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank Chie Kodama and Mizuki Morihashi for their technical assistance, Dr. Akiyoshi Fukamizu (University of Tsukuba) for his continuous support, and Zenis (https://www.zenis.co.jp) for English language editing.
Author contributions: H.M. and K.T. designed the experiments. H.M., T.T., D.K. and K.H. performed the experiments. H.M., T.T., D.K., K.H. and K.T. analyzed the data. H.M. and K.T. wrote the manuscript. All authors reviewed the results and approved the final version of the manuscript.

FUNDING

JSPS (Japan Society for the Promotion of Science) KAKENHI [JP19KK0384 to H.M., JP20K06481 to H.M., JP19H03134 to K.T., JP20K21360 to K.T., JP22H00394 to K.T.]; MEXT (Ministry of Education, Culture, Sports, Science and Technology) KAKENHI [JP20H05379 to H.M.]; Takeda Science Foundation (to K.T.). Funding for open access charge: KAKENHI grants from JSPS/MEXT.
Conflict of Interest statement. None declared.

REFERENCES

1. Barlow,D.P., Stoger,R., Herrmann,B.G., Saito,K. and Schweifer,N. (1991) The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. *Nature*, **349**, 84–87.
2. DeChiara,T.M., Robertson,E.J. and Efstratiadis,A. (1991) Parental imprinting of the mouse insulin-like growth factor II gene. *Cell*, **64**, 849–859.
3. Bartolomei,M.S., Zemel,S. and Tilghman,S.M. (1991) Parental imprinting of the mouse H19 gene. *Nature*, **351**, 153–155.
4. Plasschaert,R.N. and Bartolomei,M.S. (2014) Genomic imprinting in development, growth, behavior and stem cells. *Development*, **141**, 1805–1813.
5. Tucci,V., Isles,A.R., Kelsey,G., Ferguson-Smith,A.C. and Erice Imprinting,G. (2019) Genomic imprinting and physiological processes in mammals. *Cell*, **176**, 952–965.
6. Barlow,D.P. and Bartolomei,M.S. (2014) Genomic imprinting in mammals. *Cold Spring Harb. Perspect. Biol.*, **6**, a018382.
7. Farhadova,S., Gomez-Velazquez,M. and Feil,R. (2019) Stability and Lability of Parental Methylation Imprints in Development and Disease. *Genes (Basel)*, **10**, 999.
8. Hanna,C.W. and Kelsey,G. (2021) Features and mechanisms of canonical and noncanonical genomic imprinting. *Genes Dev.*, **35**, 821–834.
9. Kobayashi,H. (2021) Canonical and non-canonical genomic imprinting in rodents. *Front. Cell Dev. Biol.*, **9**, 713878.
10. Inoue,A., Jiang,L., Lu,F., Suzuki,T. and Zhang,Y. (2017) Maternal H3K27me3 controls DNA methylation-independent imprinting. *Nature*, **547**, 419–424.
11. Chen,Z., Yin,Q., Inoue,A., Zhang,C. and Zhang,Y. (2019) Allelic H3K27me3 to allelic DNA methylation switch maintains noncanonical imprinting in extraembryonic cells. *Sci. Adv.*, **5**, eaay7246.
12. Hanna,C.W., Perez-Palacios,R., Gahurova,L., Schubert,M., Krueger,F., Biggins,L., Andrews,S., Colome-Tatche,M., Bourc'his,D., Dean,W. *et al.* (2019) Endogenous retroviral insertions drive non-canonical imprinting in extra-embryonic tissues. *Genome Biol.*, **20**, 225.
13. Tremblay,K.D., Saam,J.R., Ingram,R.S., Tilghman,S.M. and Bartolomei,M.S. (1995) A paternal-specific methylation imprint marks the alleles of the mouse H19 gene. *Nat. Genet.*, **9**, 407–413.
14. Tremblay,K.D., Duran,K.L. and Bartolomei,M.S. (1997) A 5′ 2-kilobase-pair region of the imprinted mouse H19 gene exhibits exclusive paternal methylation throughout development. *Mol. Cell. Biol.*, **17**, 4322–4329.
15. Olek,A. and Walter,J. (1997) The pre-implantation ontogeny of the H19 methylation imprint. *Nat. Genet.*, **17**, 275–276.
16. Kato,Y., Kaneda,M., Hata,K., Kumaki,K., Hisano,M., Kohara,Y., Okano,M., Li,E., Nozaki,M. and Sasaki,H. (2007) Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. *Hum. Mol. Genet.*, **16**, 2272–2280.
17. Bell,A.C. and Felsenfeld,G. (2000) Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature*, **405**, 482–485.
18. Hark,A.T., Schoenherr,C.J., Katz,D.J., Ingram,R.S., Levorse,J.M. and Tilghman,S.M. (2000) CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature*, **405**, 486–489.
19. Kaffer,C.R., Srivastava,M., Park,K.Y., Ives,E., Hsieh,S., Batlle,J., Grinberg,A., Huang,S.P. and Pfeifer,K. (2000) A transcriptional insulator at the imprinted H19/Igf2 locus. *Genes Dev.*, **14**, 1908–1919.
20. Lleres,D., Moindrot,B., Pathak,R., Piras,V., Matelot,M., Pignard,B., Marchand,A., Poncelet,M., Perrin,A., Tellier,V. *et al.* (2019) CTCF modulates allele-specific sub-TAD organization and imprinted gene activity at the mouse Dlk1-Dio3 and Igf2-H19 domains. *Genome Biol.*, **20**, 272.
21. Matsuzaki,H., Miyajima,Y., Fukamizu,A. and Tanimoto,K. (2021) Orientation of mouse H19 ICR affects imprinted H19 gene expression through promoter methylation-dependent and -independent mechanisms. *Commun Biol*, **4**, 1410.
22. Thorvaldsen,J.L., Duran,K.L. and Bartolomei,M.S. (1998) Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2. *Genes Dev.*, **12**, 3693–3702.

23. Srivastava,M., Hsieh,S., Grinberg,A., Williams-Simons,L., Huang,S.P. and Pfeifer,K. (2000) H19 and Igf2 monoallelic expression is regulated in two distinct ways by a shared cis acting regulatory region upstream of H19. *Genes Dev.*, **14**, 1186–1195.
24. Tanimoto,K., Shimotsuma,M., Matsuzaki,H., Omori,A., Bungert,J., Engel,J.D. and Fukamizu,A. (2005) Genomic imprinting recapitulated in the human beta-globin locus. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 10250–10255.
25. Matsuzaki,H., Okamura,E., Takahashi,T., Ushiki,A., Nakamura,T., Nakano,T., Hata,K., Fukamizu,A. and Tanimoto,K. (2015) De novo DNA methylation through the 5′-segment of the H19 ICR maintains its imprint during early embryogenesis. *Development*, **142**, 3833–3844.
26. Okamura,E., Matsuzaki,H., Sakaguchi,R., Takahashi,T., Fukamizu,A. and Tanimoto,K. (2013) The H19 imprinting control region mediates preimplantation imprinted methylation of nearby sequences in yeast artificial chromosome transgenic mice. *Mol. Cell. Biol.*, **33**, 858–871.
27. Matsuzaki,H., Okamura,E., Kuramochi,D., Ushiki,A., Hirakawa,K., Fukamizu,A. and Tanimoto,K. (2018) Synthetic DNA fragments bearing ICR cis elements become differentially methylated and recapitulate genomic imprinting in transgenic mice. *Epigenetics Chromatin*, **11**, 36.
28. Matsuzaki,H., Kuramochi,D., Okamura,E., Hirakawa,K., Ushiki,A. and Tanimoto,K. (2020) Recapitulation of gametic DNA methylation and its post-fertilization maintenance with reassembled DNA elements at the mouse Igf2/H19 locus. *Epigenetics Chromatin*, **13**, 2.
29. Richard Albert,J., Au Yeung,W.K., Toriyama,K., Kobayashi,H., Hirasawa,R., Brind'Amour,J., Bogutz,A., Sasaki,H. and Lorincz,M. (2020) Maternal DNMT3A-dependent de novo methylation of the paternal genome inhibits gene expression in the early embryo. *Nat. Commun.*, **11**, 5417.
30. Quenneville,S., Verde,G., Corsinotti,A., Kapopoulou,A., Jakobsson,J., Offner,S., Baglivo,I., Pedone,P.V., Grimaldi,G., Riccio,A. *et al.* (2011) In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Mol. Cell*, **44**, 361–372.
31. Messerschmidt,D.M., de Vries,W., Ito,M., Solter,D., Ferguson-Smith,A. and Knowles,B.B. (2012) Trim28 is required for epigenetic stability during mouse oocyte to embryo transition. *Science*, **335**, 1499–1502.
32. Takahashi,N., Coluccio,A., Thorball,C.W., Planet,E., Shi,H., Offner,S., Turelli,P., Imbeault,M., Ferguson-Smith,A.C. and Trono,D. (2019) ZNF445 is a primary regulator of genomic imprinting. *Genes Dev.*, **33**, 49–54.
33. Sakaguchi,R., Okamura,E., Matsuzaki,H., Fukamizu,A. and Tanimoto,K. (2013) Sox-Oct motifs contribute to maintenance of the unmethylated H19 ICR in YAC transgenic mice. *Hum. Mol. Genet.*, **22**, 4627–4637.
34. Tanimoto,K., Liu,Q., Bungert,J. and Engel,J.D. (1999) The polyoma virus enhancer cannot substitute for DNase I core hypersensitive sites 2-4 in the human beta-globin LCR. *Nucleic Acids Res.*, **27**, 3130–3137.
35. Tanimoto,K., Liu,Q., Grosveld,F., Bungert,J. and Engel,J.D. (2000) Context-dependent EKLF responsiveness defines the developmental specificity of the human epsilon-globin gene in erythroid cells of YAC transgenic mice. *Genes Dev.*, **14**, 2778–2794.
36. de Vries,W.N., Binns,L.T., Fancher,K.S., Dean,J., Moore,R., Kemler,R. and Knowles,B.B. (2000) Expression of Cre recombinase in mouse oocytes: a means to study maternal effect genes. *Genesis*, **26**, 110–112.
37. Cong,L., Ran,F.A., Cox,D., Lin,S., Barretto,R., Habib,N., Hsu,P.D., Wu,X., Jiang,W., Marraffini,L.A. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819–823.
38. Matsuzaki,H., Okamura,E., Shimotsuma,M., Fukamizu,A. and Tanimoto,K. (2009) A randomly integrated transgenic H19 imprinting control region acquires methylation imprinting independently of its establishment in germ cells. *Mol. Cell. Biol.*, **29**, 4595–4603.
39. Chang,C., Da Silva,S.L., Ideta,R., Lee,Y., Yeh,S. and Burbach,J.P. (1994) Human and rat TR4 orphan receptors specify a subclass of the steroid receptor superfamily. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 6040–6044.
40. Glass,C.K. (1994) Differential recognition of target genes by nuclear receptor monomers, dimers, and heterodimers. *Endocr. Rev.*, **15**, 391–407.
41. Lee,C.H., Chinpaisal,C. and Wei,L.N. (1998) A novel nuclear receptor heterodimerization pathway mediated by orphan receptors TR2 and TR4. *J. Biol. Chem.*, **273**, 25209–25215.
42. Matsuzaki,H., Sugihara,S. and Tanimoto,K. (2023) The transgenic IG-DMR sequence of the mouse Dlk1-Dio3 domain acquired imprinted DNA methylation during the post-fertilization period. *Epigenetics Chromatin*, **16**, 7.
43. Saito,T., Hara,S., Kato,T., Tamano,M., Muramatsu,A., Asahara,H. and Takada,S. (2018) A tandem repeat array in IG-DMR is essential for imprinting of paternal allele at the Dlk1-Dio3 domain during embryonic development. *Hum. Mol. Genet.*, **27**, 3283–3292.
44. Watanabe,T., Tomizawa,S., Mitsuya,K., Totoki,Y., Yamamoto,Y., Kuramochi-Miyagawa,S., Iida,N., Hoki,Y., Murphy,P.J., Toyoda,A. *et al.* (2011) Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse Rasgrf1 locus. *Science*, **332**, 848–852.
45. Pearsall,R.S., Plass,C., Romano,M.A., Garrick,M.D., Shibata,H., Hayashizaki,Y. and Held,W.A. (1999) A direct repeat sequence at the Rasgrf1 locus and imprinted expression. *Genomics*, **55**, 194–201.
46. Yoon,B.J., Herman,H., Sikora,A., Smith,L.T., Plass,C. and Soloway,P.D. (2002) Regulation of DNA methylation of Rasgrf1. *Nat. Genet.*, **30**, 92–96.
47. Holmes,R., Chang,Y. and Soloway,P.D. (2006) Timing and sequence requirements defined for embryonic maintenance of imprinted DNA methylation at Rasgrf1. *Mol. Cell. Biol.*, **26**, 9564–9570.
48. Nuclear Receptors Nomenclature Committee (1999) A unified nomenclature system for the nuclear receptor superfamily. *Cell*, **97**, 161–163.
49. Cui,S., Kolodziej,K.E., Obara,N., Amaral-Psarris,A., Demmers,J., Shi,L., Engel,J.D., Grosveld,F., Strouboulis,J. and Tanabe,O. (2011) Nuclear receptors TR2 and TR4 recruit multiple epigenetic transcriptional corepressors that associate specifically with the embryonic beta-type globin promoters in differentiated adult erythroid cells. *Mol. Cell. Biol.*, **31**, 3298–3311.
50. Yang,H., Bai,D., Li,Y., Yu,Z., Wang,C., Sheng,Y., Liu,W., Gao,S. and Zhang,Y. (2022) Allele-specific H3K9me3 and DNA methylation co-marked CpG-rich regions serve as potential imprinting control regions in pre-implantation embryo. *Nat. Cell Biol.*, **24**, 783–792.