A 210-kb Segment of Tandem Repeats and Retroelements Located between Imprinted Subdomains of Mouse Distal Chromosome 7

Hisao SHIROHIZU,1 Takaaki YOKOMINE,1,2 Chiyoko SATO,3 Reiko KATO,4† Atsushi TOYODA,4
Wahyu PURBOWASITO,1,5 Chikako SUDA,1 Tsunehiro MUKAI,6 Masahira HATTORI,4,7
Katsuzumi OKUMURA,3 Yoshiyuki SAKAKI,4,8 and Hiroyuki SASAKI1,2,*

Division of Human Genetics, Department of Integrated Genetics, National Institute of Genetics, Research Organization of Information and Systems, 1111 Yata, Mishima 411-8540, Japan,1 Department of Genetics, School of Life Science, Graduate University for Advanced Studies, 1111 Yata, Mishima 411-8540, Japan,2 Laboratory of Molecular and Cellular Biology, Department of Life Science, Faculty of Bioresources, Mie University, 1515 Kamihama, Tsu 514-8507, Japan,3 Human Genome Research Group, RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan,4 Division of Disease Genes, Research Center for Genetic Information, Medical Institute of Bioregulation, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan5 Division of Molecular Biology and Genetics, Department of Biomolecular Sciences, Faculty of Medicine, Saga University, 5-1-1 Nabeshima, Saga 849-8501, Japan,6 Kitasato Institute for Life Science, Kitasato University, 1-15-1 Kitasato, Sagamihara 228-8555, Japan,7 and Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan8

(Received 14 July 2004)

Abstract

Mammalian genes subject to genomic imprinting often form clusters and are regulated by long-range mechanisms. The distal imprinted domain of mouse chromosome 7 is orthologous to the Beckwith-Wiedemann syndrome domain in human chromosome band 11p15.5 and contains at least 13 imprinted genes. This domain consists of two subdomains, which are respectively regulated by an imprinting center. We here report the finished-quality sequence of a 0.6-Mb region encompassing the more centromeric subdomain. The sequence contains four imprinted genes (Ascl2/Mash2, Ins2, Igf2 and H19) and reveals previously unidentified CpG islands and tandem repeats, which may be features of imprinted genes. Most interestingly, a unique 210-kb segment consisting almost exclusively of tandem repeats and retroelements is identified. This segment, located between Th and Ins2, has features of heterochromatin-forming DNA and is highly methylated at CpG sites. The segment exhibits asynchronous replication on the parental chromosomes, a feature of the imprinted domains. We propose that this repeat segment could serve either as a boundary between the two subdomains or as a target for epigenetic chromatin modifications that regulate imprinting.

Key words: genomic imprinting; tandem repeat; retroelement; DNA methylation; replication timing

1. Introduction

Genomic imprinting is a germline-specific epigenetic modification of the genome whereby certain mammalian genes show parental-origin-specific expression in the offspring.1–3 The number of imprinted genes seems small (up to a few hundred) in humans and mice but imprinting has a great impact on development, as exemplified by the lethality of uniparental (parthenogenetic, gynogenetic and androgenetic) embryos.4,5 Deregulation of imprinting and mutations affecting the imprinted genes can result in malformation syndromes and tumors.6 The precise mechanism governing genomic imprinting is unknown, but the experiments in mice indicated that DNA methylation is involved.7,8

A striking feature of the imprinted genes is their tendency to form clusters. For example, there is a large imprinted domain in mouse distal chromosome 7, which is orthologous to the Beckwith-Wiedemann syndrome (BWS) domain in human chromosome band 11p15.5.9 This imprinted mouse domain is about 1 Mb in size and contains at least 13 imprinted genes. Evidence in-
vides that this domain consists of at least two subdomains, each of which is regulated by an imprinting center (IC).\textsuperscript{10–13} Although the roles of long-acting enhancers and insulators (or boundary elements) are beginning to be understood,\textsuperscript{14–16} the domain-wide imprinting control mechanisms are largely unknown.

We previously identified conserved tissue-specific enhancers and methylation-sensitivity factor binding sites near the imprinted \textit{H19} gene based on a human-mouse sequence comparison.\textsuperscript{17} Towards more comprehensive studies, Onyango et al.\textsuperscript{18} reported a draft sequence of the 1-Mb domain. The finished-quality DNA sequences covering about half of the domain were then reported by two groups,\textsuperscript{19–21} but the other half spanning from \textit{Ascl2}/\textit{Mash2} to \textit{H19}, which contains a highly repetitive segment, has been left unfinished. This region encompasses the \textit{Ins2-Igf2-H19} subdomain, which is located more centromeric relative to the other subdomain.

We here report a 582-kb finished-quality sequence covering the \textit{Ins2-Igf2-H19} subdomain and some neighboring genes including \textit{Ascl2}/\textit{Mash2}. Notably, we identified a 210-kb segment composed exclusively of tandem repeats and retroelements. This segment could constitute the functional boundary between the two imprinted subdomains, or could play a role in allele-specific epigenetic chromatin modification.

\section{Materials and Methods}

\subsection{Isolation and sequencing of \textit{BAC} clones}

Isolation of mouse bacterial artificial chromosome (BAC) clones 308017, 189M11, 229F13 and 271I8 from a CITB library (mouse strain 129/SV, Research Genetics) was described previously.\textsuperscript{22} RP23-17N3 and RP23-33J1 were isolated from an RPCI-23 library (mouse strain C57BL/6J, Advanced GenoTechs)\textsuperscript{23} by PCR screening. PCR primers used were: 271I8-end sequence-tagged site (STS), 5’-TGT CCC AGG TCA TAA TGG AAC-3’ and 5’-TAG TTC CCA TTG AGA CCC TGA A-3’; 189M11-end STS, 5’-GCT GGT ACT TAC TGT GGT CTA G-3’ and 5’-GAA CCT GGG CAA GGG TTT AAG A-3’. A combined shotgun\textsuperscript{24}/nested-deletion\textsuperscript{25} strategy was adopted to sequence the BAC inserts as described previously.\textsuperscript{26} Sequence data were assembled with the Phred-Phrap and Sequencer software (Gene Codes) and deposited in DDBJ under accession numbers AP002736, AP003145, AP003146, and AP003182-AP003184.

\subsection{Sequence data analysis}

Database homology screening and gene predictions were performed using BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) and GENSCAN (http://genes.mit.edu/GENSCAN.html), respectively. CpG islands were predicted using cpgplot (http://bioinfo.pasteur.fr/sequanal/interfaces/cpgplot.html). Large-scale DNA sequence alignment was performed with Advanced PipMaker (http://nog.cse.psu.edu/pipmaker/). To eliminate spurious matches resulting solely from low and high complexity repeats, we masked the reference sequence using RepeatMasker (http://repeatmasker.genome.washington.edu/cgi-bin/RepeatMasker) before performing the PipMaker analysis. Tandem Repeats Finder (http://c3.biomath.mssm.edu/trf.html) was used to identify tandem repeats in the sequence.

\subsection{Bisulphite methylation analysis}

DNA isolated from embryos and placenta was subjected to bisulphite methylation analysis.\textsuperscript{27} The bisulphite treatment was carried out using an EZ DNA Methylation Kit (Zymo Research). Semi-nested PCR was performed to amplify the long terminal repeats (LTRs) of intracisternal A particle (IAP) sequences. The primers used were: 5’ LTR up (outer), 5’-GGA TGG ATG ATT GTT TTA ATG TGT TTT TGG AT-3’; 5’ LTR up (inner), 5’-GTT AGT AGG AAT TTT ATT GAG GA-3’; 5’ LTR down, 5’-TAT CAC TCT CTA ATT AAC TAC AAC-3’; 5’ LTR up (outer), 5’-TGG TAG GGA GAT ATG TTA TTT TGG AT-3’. PCR was carried out under the following condition: 30 cycles of 94°C for 30 sec, 55°C for 30 sec, and 72°C for 30 sec, with a final extension at 72°C for 5 min. PCR products were gel-purified using QIAquick (Qiagen), cloned using pGEM-T Easy Vector Systems (Promega), and sequenced using BigDye Terminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems) and an ABI 377 DNA Sequencer.

\subsection{FISH-based replication assay}

Replication timing was determined by a fluorescence \textit{in situ} hybridization (FISH)-based assay as described previously.\textsuperscript{28} The cells used were mouse lymphocytes prepared from the spleen and a near-diploid mouse fibroblast m5S cell line.\textsuperscript{29}

\section{Results and Discussion}

\subsection{Isolation and sequencing of mouse \textit{BAC} clones}

In a previous work, we isolated overlapping mouse BAC clones covering the imprinted domain on distal chromosome 7.\textsuperscript{22} To close the gap between \textit{Th} and \textit{Ins2}, another library (an RPCI-23 mouse BAC library) was screened by PCR using STSs at the BAC ends right next to the gap (from clones 271I8 and 189M11). As a result, a total of four clones were identified: the 271I8-end STS identified clones RP23-209N5 and RP23-334J1, while the 189M11-end STS identified RP23-209O22 and RP23-17N3. New STSs were developed following end sequencing of these clones, and subsequent STS content
analyses by PCR showed that RP23-17N3 and RP23-334J1 overlapped each other (Fig. 1), closing the gap between the two contigs.

Six BAC clones with minimal overlaps, spanning from C8d1/Tapa1 to Lsp1, were subjected to large-scale sequencing (Fig. 1) by a combined shotgun (at 8- to 10-fold redundancy)/nested-deletion (at 4-fold redundancy) strategy.26 This strategy is particularly useful in sequencing highly repetitive regions and gives accuracy over 99.995%.26 The 582,273-bp mouse sequence contained a total of 12 known genes/transcripts. They were arranged in the same order as in humans: C8d1/Tapa1, Phemx/Tssc6, Ascl2/Mash2, Th, Ins2, Igf2, Igf2as/Peg8, H19, Nctc1, Mrpl23/L23mrp, Tntn3 and Lsp1 from the telomeric to the centromeric end (Fig. 2). Our sequence contained only exons 1 and 2 of C8d1/Tapa1 and exons 6–14 of Lsp1. A gene prediction analysis using GENSCAN30 detected no more exon sequences. The overall gene arrangement and transcriptional polarities are consistent with those predicted from the draft sequence.18

3.2. G+C content and CpG islands

It has been proposed that imprinted genes tend to possess two or more CpG islands.18 The base composition and CpG frequency of our sequence were analyzed using the computer software cpgplot.31 The G+C content of the sequence was 47%, which is higher than the aver-
Table 1. CpG islands.

| Gene          | Location | Length (bp) |
|---------------|----------|-------------|
| Cd81/Tapa1    | Exon 1   | 354         |
| Ascl2/Mash2   | Upstream | 270         |
|               |          |             |
|               | Exon 1 - Intron 1 | 415         |
|               | Exon 2   | 669         |
|               | Intron 2 | 262         |
| Th ~ Ins2 intergenic | (Repeat segment) | 241         |
|               |          |             |
| Igf2          | Upstream | 432         |
|               | Exon 1   | 326         |
|               | Exon 1 - Intron 1 | 587         |
|               | Intron 1 - Exon 2 | 1004        |
|               | Exon 2   | 320         |
|               | Exon 6   | 207         |
| Mrpl23/L23mrp | Exon 1 - Intron 1 | 476         |

Table 2. Tandem repeats of relatively small size.

| Position | Unit size (bp) | Consensus sequence | Copy number |
|----------|----------------|--------------------|-------------|
| Mouse    |                |                    |             |
| mTR1     | 17371-17888    | 27                 | AGGCCCTGTGCCACCAAGGCTGAGCCG | 18 |
| mTR2     | 101388-102237   | 30                 | ACCTCCTGTGTCATCCTCCATCTATGAC | 19 |
| mTR3     | 3823-4079      | 11                 | GGGGTTATAGT | 22 |
| mTR4     | 97526-98259     | 41                 | TCTAGGGTCCTGAGAGGAGGCTATGCGTCCACCATG | 15 |

The nucleotide position numbers are according to *AP003182* and *AP003183*.

A total of 15 CpG islands were identified (Fig. 2 and Table 1) according to the criteria proposed by Gardiner-Garden and Frommer. Many of the islands were indeed associated with imprinted genes: four were associated with *Ascl2/Mash2* and six with *Igf2*. Three CpG islands, however, were newly identified within the large repeat segment found in this study (see below). While two of them corresponded to the LTRs of an IAP sequence, an endogenous retrovirus-like mobile element, the last one was in the body of this element. The imprinted *Cd81/Tapa1* gene and non-imprinted *Mrpl23/L23mrp* gene were respectively associated with one CpG island.

### 3.3. Tandem repeats associated with genes

Association with tandem repeats may be another feature of imprinted genes. At least for mouse *Impact* and *Rasgrf1*, the tandem repeats appear to play important functional roles in imprinting. Tandem Repeat Finder was used to identify such sequences. Four tandem repeats (mTR1–mTR4) with a unit size of 6 bp or more were identified (Fig. 2 and Table 2). Repeats mTR1 (upstream of *Igf2*) and mTR3 (upstream of *H19*) were previously reported. Repeat mTR2, which consisted of 19 copies of a 30-bp sequence, was located 4 kb upstream (telomeric) relative to mTR3 in the 5′ flanking region of *H19*. Repeat mTR4 was located in intron 5 of a non-imprinted gene *Lsp1* and consisted of 15 copies of a 41-bp GC-rich sequence.

### 3.4. A 210-kb segment composed of tandem repeats and retroelements

In the middle part of the 0.6-Mb sequence, we identified an extraordinarily large cluster of tandem repeats, LINE-1 retrotransposons and endogenous retrovirus-like sequences. This region (the repeat segment) was approximately 210 kb in size and located between *Th* and *Ins2*. Figure 3 shows the locations of the retroelements revealed by RepeatMasker and the locations of the tandem repeats revealed by self-sequence alignment using PipMaker. A fundamental component, occupying up to 30% of the segment, was the blocks of tandem repeat consisting basically of a 30-bp consensus sequence, 5′-AAG GCC TTA GGC CCC AGT GAG AAC CTG GAG-3′. The individual copies, however, varied greatly in both length and sequence: they contained various base substitutions and/or insertions/deletions compared with the consensus sequence. Copies with truncation or the...
Figure 3. Detailed structure of the 210-kb repeat segment between Th and Ina2. Retroelements were identified by RepeatMasker and a self-identity plot analysis was done using PipMaker. The highly repetitious tandem repeats appear as black blocks consisting of numerous dots and are marked in red. Three unrelated sequences that appear a few to several times within this region are marked in green, blue or purple.
addition of unrelated sequences were also common. Tandem repeats with this consensus sequence were not found anywhere else in the mouse genome. Another important component of the segment was the retroelements (Fig. 3). LINE-1 elements and endogenous retrovirus-like elements such as IAP were present at high densities. They occupied approximately 46% of the segment. SINE sequences were also found. The third component was 3 to 4 repetitions of three unrelated sequences (Fig. 3). The low-copy-number repeats were interspersed within the 210 kb segment, frequently interrupted by retroelements, and occupied approximately 8% of the segment. Collectively, this 210-kb segment consisted almost exclusively of three different classes of repeat sequences.

3.5. DNA methylation status at the CpG islands within the repeat segment

Tandem repeats such as satellite DNA and retroelements are often methylated at CpG sites. The heterochromatic knob of Arabidopsis, which is also composed of tandem repeats and retroelements, is also methylated.43 As noted above, three CpG islands were present (Fig. 2 and Table 1) in a full-length copy of IAP (116.7–123.8 kb in Fig. 3). We examined the two LTR CpG islands by bisulphite genomic sequencing involving PCR. To ensure specific amplification, one primer of each pair was designed for a sequence outside the IAP copy. As shown in Fig. 4, both LTR CpG islands were highly methylated in the embryo proper and also in the placenta at embryonic day 12.5 (E12.5), suggesting that this IAP copy is epigenetically silenced. This contrasts with the allele-specific 50% methylation found in the differentially methylated regions of the imprinted genes.

3.6. Asynchronous replication of the parental copies of the repeat segment

Imprinted domains are known to replicate asynchronously on the paternal and maternal chromosomes.44,45 The replication timing property along this domain, except for the repeat segment, was previously analyzed by FISH46 and S-phase fractionation.46,47 We recently isolated cosmid clones from the repeat segment.48 Two clones were selected based on their low retroelement content and used as probes in a FISH-based analysis.

As summarized in Fig. 5, the repeat probes showed a high percentage of nuclei with one single and one double dot (SD) (average 25.9%) in splenic lymphocytes, as did the other probes from the domain (average 26.9%). This suggests that the segment has differences in replication timing between the parental chromosomes. By contrast, the probes located just outside the imprinted domain and a control probe from a synchronously replicating genome region showed low percentage of such nuclei (10.8%). Essentially the same result was obtained in the nuclei of near-diploid mouse m5S fibroblasts (data not shown).
3.7. Possible role of the repeat segment

As described above, the sequence composition of the repeat segment is reminiscent of that of the heterochromatic knobs or chromomeres on the chromosome arms of plants.\textsuperscript{43,49,50} Like the heterochromatic knobs, the 210-kb segment has no active genes and is highly methylated at CpG sites. One possibility is that this candidate heterochromatin-forming segment serves as a boundary between the imprinted subdomains. Consistent with this view, we recently found that nuclear matrix attachment regions (MARs),\textsuperscript{51} which are known to function as boundaries, are clustered in this segment.\textsuperscript{48} Interestingly, between human \textit{ASCL2}/\textit{HASH2} and \textit{TH}, there is a 45-kb segment that is full of LINE-1 and retrovirus-like sequences.

In another example, retroelement-rich DNA is found at an imprinted/non-imprinted boundary. The mouse \textit{Frat3} gene is located at the telomeric boundary of another imprinted domain on mouse chromosome 7 (the chromosome 7C imprinted domain)\textsuperscript{52} and a region telomeric of this gene is extremely rich in LINE-1 sequences.\textsuperscript{52} Interestingly, the corresponding imprinted/non-imprinted transition region of the human 15q11-q13 Prader-Willi syndrome domain contains a very high density of retrovirus-like sequences but a low density of LINE-1 sequences. The observation that the separate mammalian lineages have independently accumulated retroelements at orthologous positions suggests a biological significance.

Alternatively, the repeat segment could play a more fundamental role in imprinting. A previous study showed that the functionally important IC of the imprinted human 15q11-q13 domain is composed of a high density of MARs, is associated with nuclear matrices specifically on the maternal chromosome, and is more condensed on this chromosome.\textsuperscript{53} It was proposed that this human IC is organized as heterochromatin during oogenesis and that this heterochromatin spreads along the domain during development.\textsuperscript{53} Similarly, our segment could assume different chromatin organizations in the male and female germlines, which could then persist and propagate through fertilization and development. Such an allelic difference in chromatin organization is often detected as asynchronous replication, which is indeed the feature of this repeat segment.

In either case, DNA methylation and heterochromatin formation could be induced through retroelement sequences since they are the primary targets for epigenetic silencing. A possible link between epigenetic modifications of parasitic DNA and imprinting was discussed previously.\textsuperscript{54,55} At least five imprinted genes appear to have arisen through a retrotransposition event: \textit{U2af-rs1} (a processed mRNA),\textsuperscript{56} \textit{Mkrn3} (a processed mRNA),\textsuperscript{57} \textit{Frat3} (a processed mRNA),\textsuperscript{52} \textit{Peg10} (a retrovirus)\textsuperscript{58} and \textit{Rtl1} (a retrotransposon-like sequence).\textsuperscript{59}

3.8. Conclusions

We have reported a 0.6-Mb sequence encompassing the imprinted \textit{Ins2-Igf2-H19} subdomain. Together with the published sequences,\textsuperscript{19–21} the entire imprinted domain is now covered by finished quality data. This provides the basis for identification of regulatory elements involved in imprinting and will facilitate detailed functional studies.
on the mechanisms of imprinting. The most significant finding here was the identification of a 210-kb segment surprisingly rich in tandem repeats, LINE-1 elements and retroviral insertions. We propose that this repeat segment could serve either as a boundary between the subdomains or as a target for epigenetic modifications that regulate imprinting. Although its precise function is currently unknown, it is testable in mice by using the gene knockout approach.

Acknowledgements: We thank the members of Division of Human Genetics, Department of Integrated Genetics, National Institute of Genetics, Research Organization of Information and Systems and Human Genome Research Group, RIKEN Genomic Sciences Center for technical support. This work was supported by Grant-in-Aid for Scientific Research on Priority Areas (C) “Genome Biology” from the Ministry of Education, Culture, Sports, Science and Technology of Japan and by Grants for Research on Human Genome and Gene Therapy from the Ministry of Health, Labor and Welfare of Japan.

References

1. Tilghman, S. M. 1999. The sins of the fathers and mothers: genomic imprinting in mammalian development, Cell, 96, 185–193.

2. Ferguson-Smith, A. C. and Surani, M. A. 2001, Imprinting and the epigenetic asymmetry between parental genomes, Science, 293, 1086–1089.
3. Reik, W. and Walter, J. 2001, Genomic imprinting: parental influence on the genome, *Nat. Rev. Genet.*, 2, 21–32.

4. Solter, D. 1988, Differential imprinting and expression of maternal and paternal genomes, *Annu. Rev. Genet.*, 22, 127–146.

5. Surani, M. A., Kothary, R., Allen, N. D., Singh, P. B., Fundele, R., Furguson-Smith, A. C., and Barton, S. C. 1990, Genomic imprinting and development in the mouse, *Development, Supplement*, 89–98.

6. Falls, J. G., Pullord, D. J., Wylie, A. A., and Jirtle, R. L. 1999, Genomic imprinting: implications for human disease, *Am. J. Pathol.*, 154, 635–647.

7. Li, E. 2002, Chromatin modification and epigenetic reprogramming in mammalian development, *Nat. Rev. Genet.*, 3, 662–673.

8. Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E., and Sasaki, H. 2004, Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting, *Nature*, 429, 900–903.

9. Reik, W. and Maher, E. R. 1997, Imprinting in clusters: lessons from Beckwith-Wiedemann syndrome, *Trends Genet.*, 13, 330–334.

10. Leighton, P. A., Saam, J. R., Ingram, R. S., Stewart, C. L., and Tilghman, S. M. 1995, An enhancer deletion affects both H19 and Igf2 expression, *Genes Dev.*, 9, 2079–2089.

11. Caspary, T., Cleary, M. A., Baker, C. C., Guan, X. J., and Tilghman, S. M. 1998, Multiple mechanisms regulate imprinting of the mouse distal chromosome 7 gene cluster, *Mol. Cell. Biol.*, 18, 3466–3474.

12. Horike, S., Mitsuoka, Y., Meguro, M. et al. 2000, Targeted disruption of the human LIT1 locus defines a putative imprinting control element playing an essential role in Beckwith-Wiedemann syndrome, *Hum. Mol. Genet.*, 9, 2075–2083.

13. Fitzpatrick, G. V., Soloway, P. D., and Higgins, M. J. 2002, Regional loss of imprinting and growth deficiency in mice with a targeted deletion of Kldmr1, *Nature Genet.*, 32, 426–431.

14. Bell, A. C. and Felsenfeld, G. 2000, Methylation of CTCF-dependent boundary controls imprinted expression of the Igf2 gene, *Nature*, 405, 482–485.

15. 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/Ifg2 locus, *Nature*, 405, 486–489.

16. Arney, K. L. 2003, H19 and IGF2 — enhancing the confusion? *Trends Genet.*, 19, 17–23.

17. Ishihara, K., Katano, N., Furuumi, H. et al. 2000, Comparative genomic sequencing identifies novel tissue-specific enhancers and sequence elements for methylation-sensitive factors implicated in Ifg2/H19 imprinting, *Genome Res.*, 10, 664–671.

18. Onyango, P., Miller, W., Lehoczky, J. et al. 2000, Sequence and comparative analysis of the mouse 1-megabase region orthologous to the human 11p15 imprinted domain, *Genome Res.*, 10, 1697–1710.

19. Engemann, S., Stroedicke, M., Paulsen, M. et al. 2000, Sequence and functional comparison in the Beckwith-Wiedemann region: implications for a novel imprinting center and extended imprinting, *Hum. Mol. Genet.*, 9, 2691–2706.

20. Paulsen, M., El-Maarri, O., Engemann, S. et al. 2000, Sequence conservation and variability of imprinting in the Beckwith-Wiedemann syndrome gene cluster in human and mouse, *Hum. Mol. Genet.*, 9, 1829–1841.

21. Yatsuki, H., Watanabe, H., Hattori, M. et al. 2000, Sequence-based structural features between Kvdm1 and Tap1 on mouse chromosome 7F4/F5 corresponding to the Beckwith-Wiedemann syndrome region on human 11p15.5: long stretches of unusually well conserved intronic sequences of kvdm1 between mouse and human, *DNA Res.*, 7, 195–206.

22. Kato, R., Shirohzu, H., Yokomine, T., Mizuno, S., Mukai, T., and Sasaki, H. 1999, Sequence-ready 1-Mb YAC, BAC and cosmid contigs covering the distal imprinted region of mouse chromosome 7, *DNA Res.*, 6, 401–405.

23. Chen, Y. Z., Hayashi, Y., Wu, J. G. et al. 2001, A BAC-based STS-content map spanning a 35-Mb region of human chromosome 1p35-p36, *Genomics*, 74, 55–70.

24. Fleischmann, R. D., Adams, M. D., White, O. et al. 1995, Whole-genome random sequencing and assembly of Haemophilus influenzae Rd, *Science*, 269, 496–512.

25. Hattori, M., Tsukahara, F., Furuhata, Y. et al. 1997, A novel method for making nested deletions and its application for sequencing of a 300 kb region of human APP locus, *Nucleic Acids Res.*, 25, 1802–1808.

26. Hattori, M., Fujiyama, A., Taylor, T. D. et al. 2000, The DNA sequence of human chromosome 21, *Nature*, 405, 311–319.

27. Frommer, M., McDonald, L. E., Millar, D. S. et al. 1992, A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands, *Proc. Natl. Acad. Sci. USA*, 89, 1827–1831.

28. Kagotani, K., Takebayashi, S., Kohda, A. et al. 2002, Replication timing properties within the mouse distal chromosome 7 imprinting cluster, *Biosci. Biotechnol. Biochem.*, 66, 1046–1051.

29. Sasaki, M. S. and Kodama, S. 1987, Establishment and some mutational characteristics of 3T3-like near-diploid mouse cell line, *J. Cell Physiol.*, 131, 114–122.

30. Burge, C. and Karlin, S. 1997, Prediction of complete gene structures in human genomic DNA, *J. Mol. Biol.*, 268, 78–94.

31. Larsen, F., Gundersen, G., Lopez, R. and Prydz, H. 1992, CpG islands as gene markers in the human genome, *Genomics*, 13, 1095–1107.

32. Waterston, R. H., Lindblad-Toh, K., Birney, E. et al. 2002, Initial sequencing and comparative analysis of the mouse genome, *Nature*, 420, 520–562.

33. Gardiner-Garden, M. and Frommer, M. 1987, CpG islands in vertebrate genomes, *J. Mol. Biol.*, 196, 261–282.

34. Kuff, E. L. and Lueders, K. K. 1988, The intracisternal A particle gene family: structure and functional aspects, *Adv. Cancer Res.*, 51, 183–276.

35. Zubair, M., Hilton, K., Saam, J. R., Surani, M. A., Tilghman, S. M., and Sasaki, H. 1997, Structure and expression of the mouse L23mrp gene downstream of the imprinted H19 gene: biallelic expression and lack of interaction with the H19 enhancers, *Genomics*, 45, 290–296.

36. Neumann, B., Kubicka, P., and Barlow, D. P. 1995, Char-
characteristics of imprinted genes, *Nature Genet.*, 9, 12–13.

37. Okamura, K., Hagiwara-Takeuchi, Y., Li, T. et al. 2000, Comparative genome analysis of the mouse imprinted gene *Impact* and its nonimprinted human homolog *IMPACT*: toward the structural basis for species-specific imprinting, *Genome Res.*, 10, 1878–1889.

38. Yoon, B. J., Herman, H., Sikora, A., Smith, L. T., Plass, C., and Soloway, P. D. 2002, Regulation of DNA methylation of *Rasgrf1*, *Nature Genet.*, 30, 92–96.

39. Okamura, K., Hagiwara-Takeuchi, Y., Li, T. et al. 2000, Comparative genome analysis of the mouse imprinted gene *Impact* and its nonimprinted human homolog *IMPACT*: toward the structural basis for species-specific imprinting, *Genome Res.*, 10, 1878–1889.

40. Okamura, K., Hagiwara-Takeuchi, Y., Li, T. et al. 2000, Comparative genome analysis of the mouse imprinted gene *Impact* and its nonimprinted human homolog *IMPACT*: toward the structural basis for species-specific imprinting, *Genome Res.*, 10, 1878–1889.

41. 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, *Nature Genet.*, 9, 407–413.

42. Schwartz, S., Zhang, Z., Frazer, K. A. et al. 2000, PipMaker-a web server for aligning two genomic DNA sequences, *Genome Res.*, 10, 577–586.

43. McCombie, W. R., de la Bastide, M., Habermann, K. et al. 2000, The complete sequence of a 28-kb mouse genomic region comprising the imprinted Igf2 gene, *DNA Res.*, 3, 331–335.

44. Kitsberg, D., Selig S., Brandeis, M. et al. 1993, Allelic-specific replication timing of imprinted gene regions, *Nature Genet.*, 9, 407–413.

45. Knoll, J. H., Cheng, S., and Lalande, M. 1994, Allelic specificity of DNA replication timing in the Angelman/Prader-Willi Syndrome imprinted chromosomal region, *Nature Genet.*, 6, 41–46.

46. McCombie, W. R., de la Bastide, M., Habermann, K. et al. 2000, The complete sequence of a 28-kb mouse genomic region comprising the imprinted Igf2 gene, *DNA Res.*, 3, 331–335.

47. McCombie, W. R., de la Bastide, M., Habermann, K. et al. 2000, The complete sequence of a 28-kb mouse genomic region comprising the imprinted Igf2 gene, *DNA Res.*, 3, 331–335.

48. McCombie, W. R., de la Bastide, M., Habermann, K. et al. 2000, The complete sequence of a 28-kb mouse genomic region comprising the imprinted Igf2 gene, *DNA Res.*, 3, 331–335.

49. McCombie, W. R., de la Bastide, M., Habermann, K. et al. 2000, The complete sequence of a 28-kb mouse genomic region comprising the imprinted Igf2 gene, *DNA Res.*, 3, 331–335.

50. McCombie, W. R., de la Bastide, M., Habermann, K. et al. 2000, The complete sequence of a 28-kb mouse genomic region comprising the imprinted Igf2 gene, *DNA Res.*, 3, 331–335.

51. Kitsberg, D., Selig S., Brandeis, M. et al. 1993, Allelic-specific replication timing of imprinted gene regions, *Nature Genet.*, 9, 407–413.

52. Cockerill, P. N. and Garrard, W. T. 1986, Chromosomal loop anchorage of the kappa immunoglobulin gene occurs next to enhancer in a region containing topoisomerase II sites, *Cell*, 44, 273–282.

53. Chai, J. H., Locke, D. P., Ohta, T., Greally, J. M., and Nicholls, R. D. 2001, Retrotransposed genes such as Frat3 in the mouse Chromosome 7C Prader-Willi syndrome region acquire the imprinted status of their insertion site, *Mamm. Genome*, 12, 813–21.

54. Greally, J. M., Gray, T. A., Gabriel, J. M., Song, L., Zemel, S., and Nicholls, R. D. 1999, Conserved characteristics of heterochromatin-forming DNA at the 15q11-q13 imprinting center, *Proc. Natl. Acad. Sci. USA*, 96, 14430–14435.

55. Barlow, D. P. 1993, Methylation and imprinting: from host defense to gene regulation, *Science*, 260, 309–310.

56. McCombie, W. R., de la Bastide, M., Habermann, K. et al. 2000, The complete sequence of a 28-kb mouse genomic region comprising the imprinted Igf2 gene, *DNA Res.*, 3, 331–335.

57. Jong, M. T., Carey, A. H., Caldwell, K. A. et al. 1999, Asynchronous replication timing of imprinted gene regions, *Nature Genet.*, 9, 407–413.

58. McCombie, W. R., de la Bastide, M., Habermann, K. et al. 2000, The complete sequence of a 28-kb mouse genomic region comprising the imprinted Igf2 gene, *DNA Res.*, 3, 331–335.

59. Seitz, H., Youngson, N., Lin, S.-P. et al. 2003, Imprinted Rasgrf1 gene, *Mamm. Genome*, 14, 813–21.

60. Nabetani, A., Hatada, I., Morisaki, H., Oshimura, M., and Mukai, T. 1997, Mouse *U2af1-rs1* is a neomorphic imprinting gene, *Mol. Cell. Biol.*, 17, 789–798.

61. Jong, M. T., Carey, A. H., Caldwell, K. A. et al. 1999, Asynchronous replication timing of imprinted gene regions, *Nature Genet.*, 9, 407–413.

62. McCombie, W. R., de la Bastide, M., Habermann, K. et al. 2000, The complete sequence of a 28-kb mouse genomic region comprising the imprinted Igf2 gene, *DNA Res.*, 3, 331–335.

63. McCombie, W. R., de la Bastide, M., Habermann, K. et al. 2000, The complete sequence of a 28-kb mouse genomic region comprising the imprinted Igf2 gene, *DNA Res.*, 3, 331–335.

64. McCombie, W. R., de la Bastide, M., Habermann, K. et al. 2000, The complete sequence of a 28-kb mouse genomic region comprising the imprinted Igf2 gene, *DNA Res.*, 3, 331–335.

65. McCombie, W. R., de la Bastide, M., Habermann, K. et al. 2000, The complete sequence of a 28-kb mouse genomic region comprising the imprinted Igf2 gene, *DNA Res.*, 3, 331–335.

66. Nabetani, A., Hatada, I., Morisaki, H., Oshimura, M., and Mukai, T. 1997, Mouse *U2af1-rs1* is a neomorphic imprinting gene, *Mol. Cell. Biol.*, 17, 789–798.

67. Jong, M. T., Carey, A. H., Caldwell, K. A. et al. 1999, Asynchronous replication timing of imprinted gene regions, *Nature Genet.*, 9, 407–413.

68. McCombie, W. R., de la Bastide, M., Habermann, K. et al. 2000, The complete sequence of a 28-kb mouse genomic region comprising the imprinted Igf2 gene, *DNA Res.*, 3, 331–335.

69. McCombie, W. R., de la Bastide, M., Habermann, K. et al. 2000, The complete sequence of a 28-kb mouse genomic region comprising the imprinted Igf2 gene, *DNA Res.*, 3, 331–335.

70. Jong, M. T., Carey, A. H., Caldwell, K. A. et al. 1999, Asynchronous replication timing of imprinted gene regions, *Nature Genet.*, 9, 407–413.

71. McCombie, W. R., de la Bastide, M., Habermann, K. et al. 2000, The complete sequence of a 28-kb mouse genomic region comprising the imprinted Igf2 gene, *DNA Res.*, 3, 331–335.

72. McCombie, W. R., de la Bastide, M., Habermann, K. et al. 2000, The complete sequence of a 28-kb mouse genomic region comprising the imprinted Igf2 gene, *DNA Res.*, 3, 331–335.