Structural and functional analysis of a 0.5-Mb chicken region orthologous to the imprinted mammalian *Ascl2/Mash2–Igf2–H19* region

Takaaki Yokomine,[1,2] Hisao Shirohzu,[1] Wahyu Purbowasito,[1,3] Atsushi Toyoda,[4] Hisakazu Iwama,[5] Kazuho Ikeo,[5] Tetsuya Hori,[6,12] Shigeki Mizuno,[7] Masaoki Tsudzuki,[8] Yoh-ichi Matsuda,[9] Masahira Hattori,[4,10] Yoshiyuki Sakaki,[4,11] and Hiroyuki Sasaki[1,2,13]

[1]*Division of Human Genetics, Department of Integrated Genetics, National Institute of Genetics, Research Organization of Information and Systems and* [2]*Department of Genetics, School of Life Science, Graduate University for Advanced Studies, Mishima 411-8540, Japan;* [3]*Division of Disease Genes, Research Center for Genetic Information, Medical Institute of Bioregulation, Kyushu University, Fukuoka 812-8582, Japan;* [4]*Human Genome Research Group, Genomic Sciences Center, RIKEN Yokohama Institute, Yokohama 230-0045, Japan;* [5]*DNA Data Analysis Laboratory, Center for Information and Biology and DNA Data Bank of Japan, National Institute of Genetics, Research Organization of Information and Systems, Mishima 411-8540, Japan;* [6]*Department of Genetic Engineering, School of Biology-Oriented Science and Technology, Kinki University, Uchita, Wakayama 649-6493, Japan;* [7]*Department of Agricultural and Biological Chemistry, College of Bioresource Sciences, Nihon University, Fujisawa 252-8510, Japan;* [8]*Laboratory of Animal Genetics, Faculty of Applied Biological Science, Hiroshima University, Higashi-Hiroshima 739-8528, Japan;* [9]*Chromosome Research Unit, Faculty of Science, and Laboratory of Cytogenetics, Graduate School of Environmental Earth Science, Hokkaido University, Sapporo 060-0810, Japan;* [10]*Kitasato Institute for Life Science, Kitasato University, Sagamihara 228-8555, Japan;* [11]*Human Genome Center, Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo 108-8639, Japan*

Previous studies revealed that *Igf2* and *Mpr/Igf2r* are imprinted in eutherian mammals and marsupials but not in monotremes or birds. *Igf2* lies in a large imprinted cluster in eutherians, and its imprinting is regulated by long-range mechanisms. As a step to understand how the imprinted cluster evolved, we have determined a 490-kb chicken sequence containing the orthologs of mammalian *Ascl2/Mash2*, *Ins2* and *Igf2*. We found that most of the genes in this region are conserved between chickens and mammals, maintaining the same transcriptional polarities and exon–intron structures. However, *H19*, an imprinted noncoding transcript, was absent from the chicken sequence. Chicken *ASCL2/CASH4* and *INS*, the orthologs of the imprinted mammalian genes, showed biallelic expression, further supporting the notion that imprinting evolved after the divergence of mammals and birds. The *H19* imprinting center and many of the local regulatory elements identified in mammals were not found in chickens. Also, a large segment of tandem repeats and retroelements identified between the two imprinted subdomains in mice was not found in chickens. Our findings show that the imprinted genes were clustered before the emergence of imprinting and that the elements associated with imprinting probably evolved after the divergence of mammals and birds.

[The sequence data described in this paper have been deposited in the DDBJ/GenBank/EMBL data library under accession nos. AP003795, AP003796, AP004717, AB101638, AB101639, AB101640, and AB101641.]

Genomic imprinting, a germ-line-specific epigenetic modification of the genome, causes parental-origin-specific expression of a small subset of genes (up to a few hundred) in eutherian mammals such as humans and mice (Tilghman 1999; Ferguson-Smith and Surani 2001; Reik and Walter 2001). The imprinted genes play crucial roles in embryonic development and growth regulation (Solter 1988; Surani et al. 1990; Cattanach and Beechey 1997). They also affect behavior and several disease phenotypes (Falls et al. 1999). However, the reasons for the evolution of imprinting is not well understood. Also, although CpG methylation is clearly involved in the imprinting processes (Li 2002; Kaneda et al. 2004), the precise mechanisms governing imprinting are yet to be elucidated.

Recent studies revealed that *Igf2*, a paternally expressed gene in eutherians, shows the same imprinted expression pattern in marsupials (such as opossums) (O'Neill et al. 2000). However, this gene is biallelically expressed in monotremes (such as platypuses and echidnas) and birds (such as chickens) (O'Neill et al. 2000; Nolan et al. 2001; Yokomine et al. 2001). Similarly, it was shown that *Mpr/Igf2r*, a gene located on a different chromosome, is maternally expressed in both eutherians and marsupials, but biallelically expressed in monotremes and birds (Killian et al. 2000; Nolan et al. 2001; Yokomine et al. 2001). These results are consistent with the conflict hypothesis of imprinting evolution

[12]Present address: Division of Molecular Biology, Department of Molecular Biology, National Institute of Genetics, Research Organization of Information and Systems, Mishima 411-8540, Japan.
[13]Corresponding author.
E-mail hisasaki@lab.nig.ac.jp; fax 81-(0)55-981-6800.
Article and publication are at http://www.genome.org/cgi/doi/10.1101/gr.2609605. Article published online before print in December 2004.

15:154–165 ©2005 by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/05; www.genome.org

(Moore and Haig 1991), which predicts that imprinting would be favored by viviparous animals.

A striking feature of the imprinted genes is their tendency to form clusters, which is most probably associated with the mechanisms of imprinting. *Igf2* is contained in a large imprinted gene cluster (imprinted domain) on Chromosome 7 in mice and Chromosome 11 in humans (Reik and Maher 1997). The human domain is responsible for Beckwith-Wiedemann syndrome (BWS), an imprinting-associated fetal overgrowth syndrome. The domain is ~1 Mb in size and contains 13 imprinted genes. Evidence indicates that the domain is composed of two subdomains, which are, respectively, regulated by an imprinting center (IC) (Leighton et al. 1995b; Caspary et al. 1998; Horike et al. 2000; Fitzpatrick et al. 2002). The sequence elements in the ICs and many local regulatory elements involved in the allele-specific expression of the genes are conserved between humans and mice (Ainscough et al. 2000; Bell and Felsenfeld 2000; Hark et al. 2000; Ishihara et al. 2000; Drewell et al. 2002).

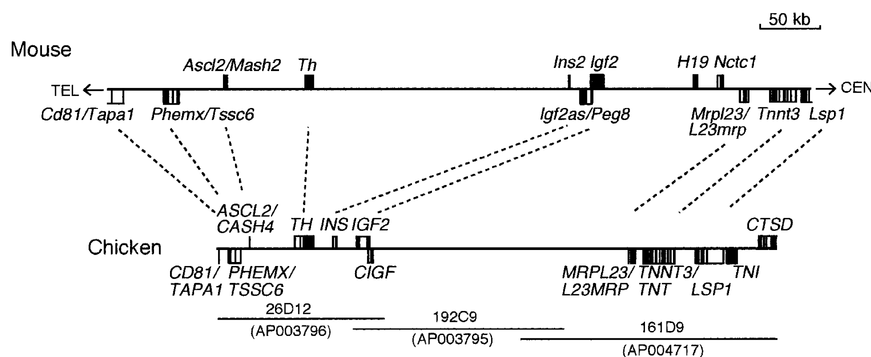

Figure 1. Overview of the *ASCL2/CASH4-IGF2* region on chicken Chromosome 5 and the orthologous region on mouse Chromosome 7. The locations of the chicken BAC clones that were sequenced in this study are shown as horizontal bars. The DDBJ/GenBank/EMBL accession numbers of the sequences are indicated in parentheses.

In order to understand how the long-range imprinting mechanisms evolved during mammalian evolution, it is important to know the structural and functional properties of the orthologous region of nonimprinted vertebrate species. Chickens provide an excellent model for such a comparative study because they have been an important experimental system in many fields of biology including developmental biology. In addition to the easy access to the embryo, the increase in genomic resources is enabling chicken research to contribute to the functional analysis of the vertebrate genome (Brown et al. 2003).

Here we report the DNA sequence of a 0.5-Mb chicken region containing *IGF2*. This enables us for the first time to make large-scale structural and functional comparisons of an imprinted mammalian region with the orthologous region from an oviparous vertebrate species. We show that the chicken orthologs of *Ascl2/Mash2* and *Ins2* are not imprinted, supporting the idea that birds do not have genomic imprinting. We also show that most of the elements involved in imprinting in humans and mice are not present in chickens. The present work provides the basis to study and understand how an imprinted region evolved and how imprinting is regulated.

Results

Isolation and sequencing of chicken BAC clones

To isolate bacterial artificial chromosome (BAC) clones containing the chicken region orthologous to the imprinted *Ascl2/Mash2–Igf2–H19* region, a White Leghorn BAC library (Hori et al. 2000) was screened with a PCR-amplified *IGF2* probe (Yokomine et al. 2001). A total of six BAC clones were obtained (26D12, 90B1, 192C9, 283C3, 411E9, and 457F4). Based on the data from sequence-tagged site (STS) content analysis and restriction fingerprinting, the clones 26D12 and 192C9 were selected for large-

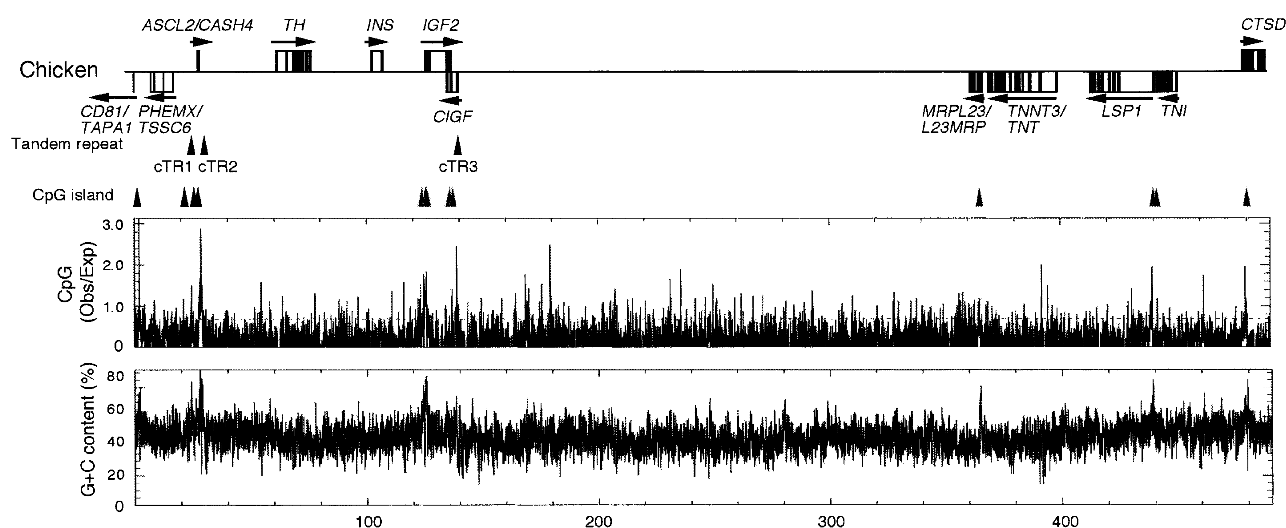

Figure 2. Overall structure and gene arrangement of the chicken region. Arrangement of genes (boxes) is illustrated at the *top* of each panel. Horizontal arrows or arrowheads *above* or *below* the genes indicate directions of transcription. Upward arrowheads indicate positions of CpG islands and tandem repeats. Observed/Expected ratio of CpG dinucleotide per 100 bp and G+C content per 100 bp are graphically shown *under* the gene arrangement.

scale sequencing (Fig. 1). To isolate clones extending to more 3′, an end probe was produced from clone 192C9 and used to rescreen the BAC library. Two clones were obtained (27B1 and 161D9). The clone 161D9 was subjected to sequencing because this clone contained the *TNNT3/TNT* (the ortholog for mouse *Tnnt3*) marker and thus should be longer than the other.

The chicken sequence of 490,074 bp contained eight previously reported genes/transcripts, *ASCL2/CASH4*, *TH*, *INS*, *IGF2*, *CIGF*, *TNNT3/TNT*, *TNI*, and *CTSD* (Fig. 2). Among these, *TH*, *INS*, *IGF2*, *TNNT3/TNT*, *TNI*, and *CTSD* were known to be the orthologs of mouse *Th*, *Ins2*, *Igf2*, *Tnnt3*, *Tni*, and *Ctsd*, respectively. *CIGF* was previously identified as an antisense transcript at the *IGF2* locus (Taylor et al. 1991). *ASCL2/CASH4*, a member of the *achaete–scute* gene family (Henrique et al. 1997), was now revealed to be the ortholog of mouse *Ascl2/Mash2*, a gene essential for placental development (Guillemot et al. 1994). In addition to these eight genes/transcripts, four genes orthologous to mouse *Cd81/Tapa1*, *Phemx/Tssc6*, *Mrpl23/L23mrp*, and *Lsp1*, respectively, were identified by data analysis (see below) (Fig. 2).

We recently reported a 0.6-Mb mouse sequence corresponding to this sequence (Shirohzu et al. 2004). Comparisons of the two sequences revealed that the gene order and transcriptional polarities are conserved except that two genes/transcripts are missing in chickens (see below). The gene distance, however, varied greatly between the two species. For example, a chicken region spanning from *CD81/TAPA1* to *IGF2* (~140 kb) was three times smaller than the orthologous region in mice (Figs. 1 and 2). The more compact organization of the chicken region was due to shorter intergenic distances. The absence of the 210-kb region composed of tandem repeats and retroelements, which was found between *Th* and *Ins2* in mice (Shirohzu et al. 2004), was another reason for the more compact structure. In contrast, the intergenic region between chicken *IGF2* and *MRPL23/L23MRP* (~223 kb) was twice as large as the corresponding region in mice.

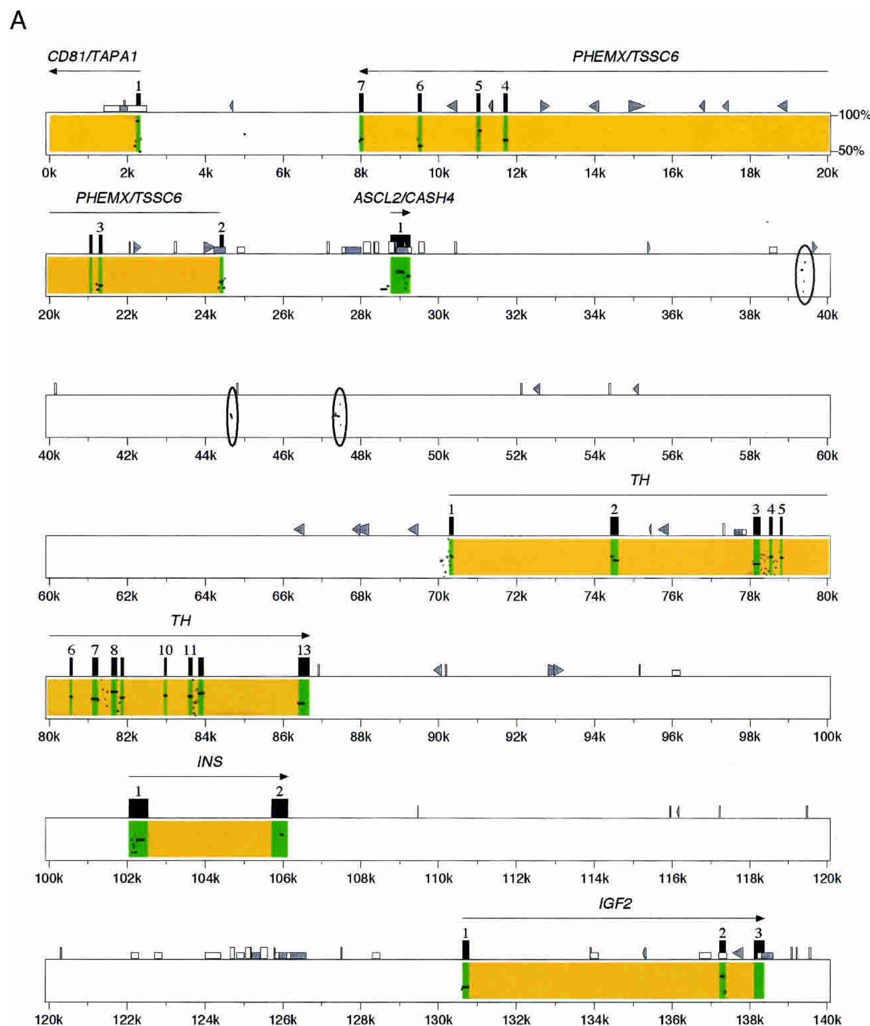

Figure 3. (Continued on next page)

Identification and characterization of chicken *CD81/TAPA1*, *PHEMX/TSSC6*, *MRPL23/L23MRP*, and *LSP1*

Chicken *CD81/TAPA1*, *PHEMX/TSSC6*, *MRPL23/L23MRP*, and *LSP1* were identified by making use of BLAST, GENSCAN, and a large-scale sequence comparison program, PipMaker (Schwartz et al. 2000). Figure 3 shows the results of PipMaker analyses, which were particularly useful in detecting the new genes. The percent identity plot studies highlighted the well-conserved exon regions in the chicken sequence.

One of the genes newly identified was *CD81/TAPA1*, a gene encoding a member of the transmembrane-4 superfamily. A 105-bp region located at ~2.2 kb from the 5′-end of our chicken sequence showed a 78.2% similarity with the exon 1 sequence of mouse *Cd81/Tapa1*. This chicken exon contained an initiation codon and an open reading frame, which codes for the first 22 amino acids (MGVEGCTKCIKYLLFVFNFVFN) of the presumed chicken CD81/TAPA1 protein (GenBank accession no. AB101638). The amino acid sequence was identical with that of the N-terminal part of human CD81/TAPA1, but there was one amino acid difference (20V to I) when compared with the mouse protein. We then examined the tissue distribution of *CD81/TAPA1* transcripts by RT-PCR and found that it is expressed ubiquitously just as in mice (Andria et al. 1991; data not shown).

Six evolutionarily conserved DNA segments were identified in the 8–25-kb region of the chicken sequence (Fig. 3). The segments were 54.8%–64.9% identical in sequence with exons 2, 3, 4, 5, 6, and 7 of mouse *Phemx/Tssc6*, a transcript with potential tumor suppressor activity (Paulsen et al. 2000). We then examined the presence of the transcripts in chicken tissues by RT-PCR using two primer pairs. With the primers for exons 2 and 7, four PCR products were detected in all embryonic stages and adult tissues that we examined (Fig. 4A). The largest and most abundant product (isoform 1) contained all the predicted exons (GenBank accession no. AB101639) (Fig. 4B). This transcript species had a single open reading frame, and its deduced amino acid

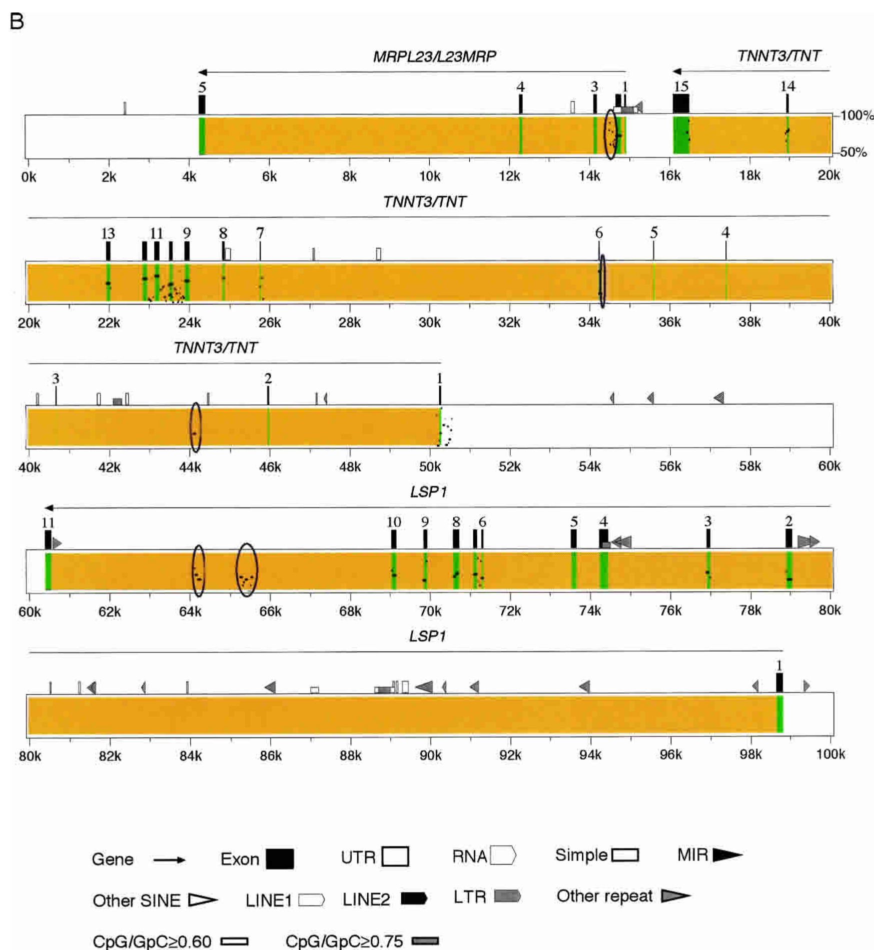

Figure 3. Sequence comparison between chickens and mice from *CD81/TAPA1* to *IGF2* (*A*) and from *MRPL23/L23MRP* to *LSP1* (*B*). Percent identity plot analysis was done using the PipMaker software using the chicken sequence as a reference. The order and arrangement of the chicken genes are shown at the *top*. The dots and short horizontal bars correspond to the segments of sequence conservation. The regions corresponding to the exons and introns of the chicken genes are shown in green and yellow, respectively. Conserved sequences in the noncoding regions, including those in introns, are circled.

sequence exhibited 45.6% and 42.2% identity with the human and the mouse PHEMX/TSSC6 protein, respectively (Fig. 4C). The second largest product (isoform 2) lacked the exon 6 sequence, the third one (isoform 3) lacked exons 4 and 6, and the smallest one (isoform 4) lacked exons 4, 5, and 6 (Fig. 4A). None of these alternative-splicing events would cause a frameshift (Fig. 4B). With the primers for exons 3 and 6, two other alternative splicing products were identified (isoforms 5 and 6). Both of them lacked exon 5, but one (isoform 5) contained an additional 68-bp sequence from intron 3 (exon 3′) (Fig. 4A). This insertion was predicted to cause a frameshift resulting in aberrant amino acids and stop codons within the exon 4 region. Although multiple splicing variants have also been observed for mouse *Phemx/Tssc6* (Paulsen et al. 2000), the one corresponding to isoform 5 has not been reported.

The chicken ortholog of mammalian *Mrpl23/L23mrp*, a gene encoding a putative mitochondrial ribosomal protein (Tsang et al. 1995; Zubair et al. 1997), was located downstream of *TNNT3/TNT*. This gene consisted of five exons as its mammalian ortholog. The predicted protein product of the chicken *MRPL23/L23MRP* cDNA (GenBank accession no. AB101640) was 154 amino acids in size and displayed 58.0% and 54.0% identity with the human and the mouse ortholog, respectively (Fig. 5A,B). RT-PCR analyses showed that the chicken gene is ubiquitously expressed just as its mammalian orthologs (Tsang et al. 1995; Zubair et al. 1997; data not shown).

Chicken *LSP1* was identified upstream of *TNNT3/TNT* at a position orthologous to mammalian *Lsp1*. Mouse *Lsp1* codes for a lymphocyte-specific calcium-binding protein with unknown function (Jongstra et al. 1988) and was previously mapped upstream of *Tnnt3* (Misener et al. 1998). The chicken gene consisted of 11 exons as the mammalian ortholog. The 5′-end of the gene was tentatively assigned based on an expressed sequence tag (EST) sequence (riken1 8h20r1), which was assumed to be a full-length cDNA, from the BursaEst Database (http://swallow.gsf.de/dt40est.html). Exons 10 and 11, which code for the 3′-untranslated sequences, were predicted based on another EST sequence (GenBank accession no. AI979962). RT-PCR primers were designed according to these EST sequences, and the middle part of the cDNA was amplified and sequenced. The predicted protein product of the chicken *LSP1* cDNA (GenBank accession no. AB101641) was 318 amino acids in size and displayed 34.8% and 36.5% identity with the human and the mouse ortholog, respectively (Fig. 6A,B). RT-PCR analyses revealed that chicken *LSP1* is highly expressed in lymphoid tissues such as the spleen (data not shown). Expression was also detectable, however, at a comparable level in the ovary and at a lower level in the lung.

Chicken *ASCL2/CASH4* and *INS* are not imprinted

Although it has been shown that chicken *IGF2* is not imprinted (O'Neill et al. 2000; Nolan et al. 2001; Yokomine et al. 2001), the imprinting status of the other genes in this region is unknown. *ASCL2/CASH4* is of special interest because its mouse ortholog *Ascl2/Mash2* is imprinted to be silent on the maternal chromosome (Guillemot et al. 1995). However, its human ortholog *ASCL2/HASH2* does not appear to be imprinted (Miyamoto et al. 2002). *INS* is another interesting gene because its mouse and human orthologs are imprinted in a tissue-specific way: they are expressed from the paternal allele in the yolk sac but expressed from both alleles in the pancreas (Giddings et al. 1994; Moore et al. 2001).

To study the allelic expression status of *ASCL2/CASH4* and *INS*, as well as *PHEMX/TSSC6* and *TH*, of which orthologs are not imprinted in mammals, we looked for single nucleotide polymorphisms (SNPs) in their transcribed regions that can be used to trace parental origin of the alleles. Comparisons of the PCR-amplified genomic sequences from six different chicken lines

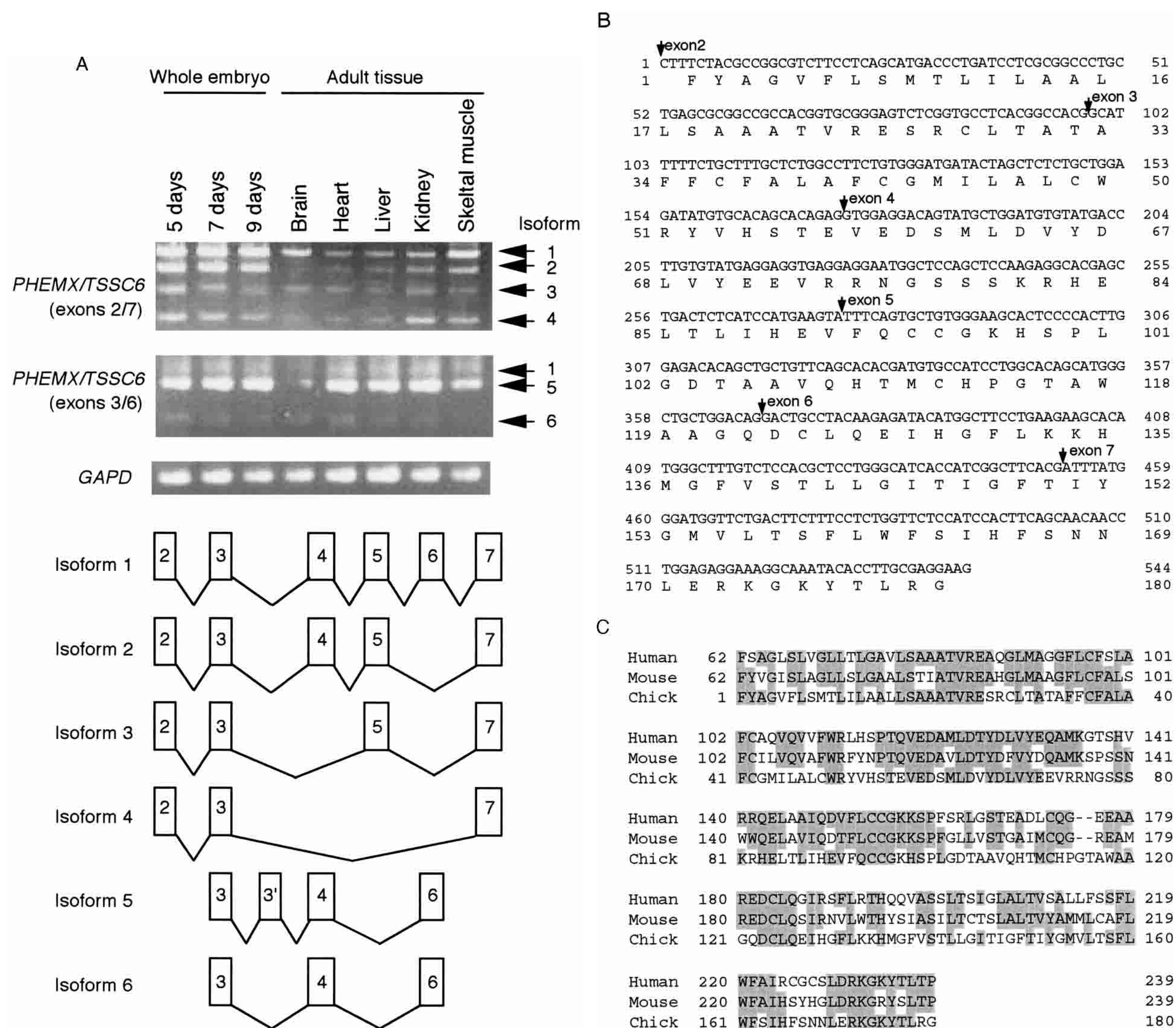

Figure 4. Structure and expression of chicken *PHEMX/TSSC6*. (*A*) RT-PCR analysis of chicken *PHEMX/TSSC6* in whole chick embryos and adult tissues. *GAPD* was used as a housekeeping control. Splicing variants, revealed by sequencing of the PCR products, are shown (isoforms 1–6). (*B*) Nucleotide sequence and predicted amino acid sequence of chicken *PHEMX/TSSC6* cDNA (isoform 1; GenBank accession no. AB101639). Arrowheads indicate the positions of exon–intron boundaries. (*C*) Alignment of the predicted amino acid sequences of human (GenBank accession no. AF125569), mouse (GenBank accession no. AJ279791), and chicken *PHEMX/TSSC6*. Amino acids conserved in more than two species are shaded.

uncovered at least three SNPs for each gene (Table 1). The chicken lines with different SNP genotypes were reciprocally crossed, and the RT-PCR products from resulting embryos were sequenced. We first confirmed the biallelic expression of *IGF2* in four informative embryos recovered at day 5 of development (5-d embryos) using the SNP identified previously (Fig. 7A; Yokomine et al. 2001). We then found that *ASCL2/CASH4* is biallelically expressed in nine 1-d embryos (stage 5–6) and in the embryonic membranes (chorion, allantoic membrane, and yolk sac) of 14 5-d embryos (Fig. 7B). We also demonstrated biallelic expression of *INS*, *PHEMX/TSSC6*, and *TH* in all informative 5-d embryos (6, 11, and 8 embryos, respectively) (Fig. 7C–E). Biallelic expression of *INS* was also detected in the embryonic membranes, including the yolk sac, of nine 5-d embryos (Fig. 7C). Thus, all genes examined here, including *ASCL2/CASH4* and *INS*, were not imprinted, although we cannot exclude the possibility that they are imprinted in limited tissues of the embryos or at other developmental stages.

Absence of *H19* and *Nctc1* in chickens

In eutherian mammals, a maternally expressed imprinted transcript, *H19*, is located between *Igf2* and *Mrpl23/L23mrp*, at about one-third of the way from *Mrpl23/L23mrp* (Figs. 1 and 2). *H19* does not code for a protein but is expressed at high levels in a wide array of mesodermal and endodermal tissues. Approximately 17 kb 3′ to mouse *H19*, another noncoding transcript called *Nctc1* is present (Ishihara et al. 1998). *Nctc1* is expressed in the skeletal muscle and shows biallelic expression. No sequence orthologous to *H19* or *Nctc1* was found in the chicken sequence.

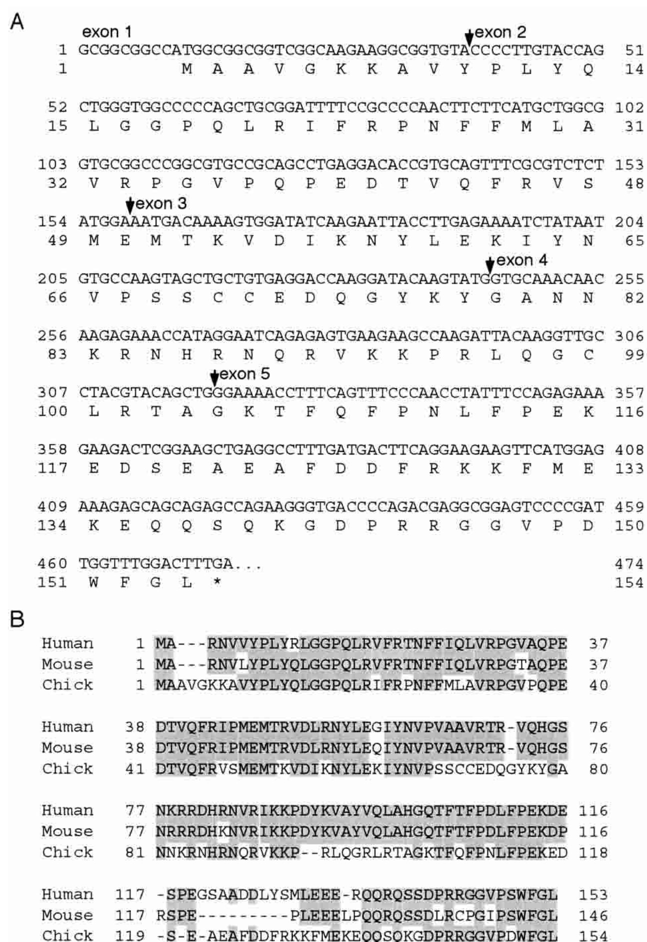

Figure 5. Structure of chicken *MRPL23/L23MRP*. (*A*) Nucleotide sequence and predicted amino acid sequence of chicken *MRPL23/L23MRP* cDNA (GenBank accession no. AB101640). Arrowheads indicate the positions of exon–intron boundaries. (*B*) Alignment of the predicted amino acid sequences of human (GenBank accession no. BC027710), mouse (GenBank accession no. U84902), and chicken *MRPL23/L23MRP*.

In fact, no transcript or exon-like sequence was found in the 223-kb chicken region between *IGF2* and *MRPL23/L23MRP* by BLAST search for ESTs or GENSCAN search for exons. While *Nctc1* is not conserved even in mammals (e.g., humans do not have this transcript), the existence of chicken *H19* was previously suggested by a zoo blot analysis (Brannan et al. 1990). The possibility that chickens have *H19* somewhere else in the genome remains to be investigated.

G+C content and CpG islands

It has been proposed that the imprinted genes of this domain tend to possess two or more CpG islands upstream of and/or within the gene (Onyango et al. 2000). The base composition and CpG frequency of our chicken sequence were analyzed by the computer software cpgplot (Larsen et al. 1992). The G+C content was not much different from that of the corresponding mouse sequence (46.5% vs. 47.1%), both of which were above the average G+C content of the vertebrate genomes (~40%). Thirteen CpG islands were identified in the chicken sequence (Fig. 2; Table 2), according to the criteria by Gardiner-Garden and Frommer (1987). Among these, three were associated with *ASCL2/CASH4* and five with *IGF2*. Thus, the multiple CpG island rule may also apply to the chicken orthologs, and it is clear that this is not sufficient for imprinting.

Tandem repeats

Association with tandem repeats has been suggested to be a feature of mammalian imprinted genes (Neumann et al. 1995). Indeed, tandem repeats have been shown to be associated with functional imprinting at the mouse *Impact* and *Rasgrf1* loci (Okamura et al. 2000; Yoon et al. 2002).

A program called Tandem Repeat Finder (Benson 1999) was used to identify tandem repeats. We identified three tandem repeats (cTR1–cTR3) with a unit size of 6 bp or more (Fig. 2; Table 3). The repeat cTR3, a 14-time repetition of a 18-bp sequence, was located in the 3′-flanking region of *IGF2*. The repeats cTR1 and cTR2 were located in the 5′-flanking region and the 3′ untranslated region, respectively, of *ASCL2/CASH4*: The repeat cTR1 was a 26-time repetition of a 16-bp sequence, and cTR2 was a 29-time repetition of a 6-bp sequence (Table 3). Thus, although repeats identical or very similar to cTR1–cTR3 were not present in mice, the orthologs of imprinted genes tended to possess tandem repeats.

We recently reported that a 210-kb region between mouse *Th* and *Ins2* contains numerous tandem repeats, which could serve either as a boundary between the two imprinted subdomains or as a target for epigenetic chromatin modifications leading to imprinting (Shirohzu et al. 2004). No such tandem repeat was detected in our chicken sequence.

The IC and differentially methylated regions (DMRs)

The key regulatory elements involved in imprinting are likely to be conserved among imprinted species but may be absent from nonimprinted species. Our results that the chicken genes of the region are not imprinted provided a unique opportunity to investigate this possibility.

Previous studies showed that a 2.0-kb differentially methylated region (DMR) located 1.2 kb upstream of *H19* serves as an IC: it is essential for both silencing of paternal *H19* and silencing of

Table 1. DNA polymorphisms

Gene	Location	Polymorphism (position[a])	Line or breed[b]
PHEMX/TSSC6	Exon 4	A→G (11,747)	BPR, WL-D, WL-G, WL-HA
	Exon 3′	T→C (21,082)	BPR, WL-HA
		C→T (21,098)	OS
ASCL2/CASH4	Exon 1	C→T (29,316)	BPR, WL-Cornell-P, WL-G
		A→G (29,365)	BPR, WL-Cornell-P, WL-G
		C→T (29,406)	BPR, WL-Cornell-P, WL-G
		G→A (29,429)	BPR, OS
		A→G (29,450)	BPR, OS, WL-Cornell-P, WL-G
		C→T (29,534)	BPR, WL-G
		T→C (29,563)	BPR, WL-Cornell-P, WL-G
		T→C (29,570)	BPR, OS
TH	Exon 13	C→T (86,656)	BPR
		G→T (86,744)	BPR, OS
		G→A (86,848)	WL-Cornell-P
INS	Exon 2	C→T (106,014)	OS
		G→A (106,015)	BPR, WL-Cornell-P, WL-HA
		G→T (106,047)	OS

[a]The nucleotide position number is according to AP003796.
[b](BRP), Barred Plymouth Rock; (OS), Oh-Shamo; (WL-Cornell-P), Cornell-P line of White Leghorn; (WL-D), D line of White Leghorn; (WL-G) G line of White Leghorn; (WL-HA), HA line of White Leghorn.

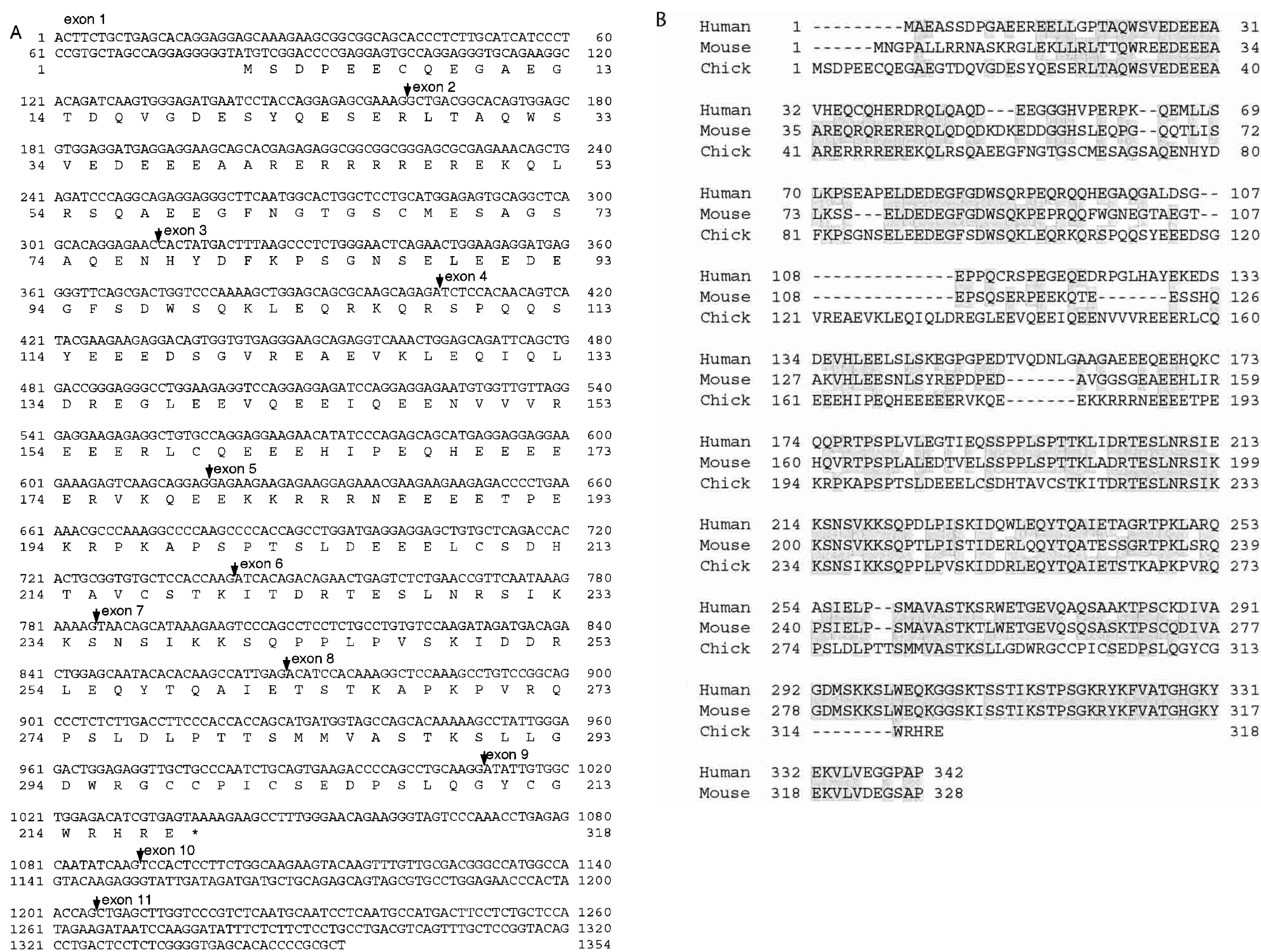

Figure 6. Structure of chicken *LSP1*. (*A*) Nucleotide sequence and predicted amino acid sequence of chicken *LSP1* cDNA (GenBank accession no. AB101641). Arrowheads indicate the positions of exon–intron boundaries. (*B*) Alignment of the predicted amino acid sequences of human (GenBank accession no. NM002339), mouse (GenBank accession no. NM019391), and chicken *LSP1*.

maternal *Igf2* (Thorvaldsen et al. 1998). This IC contains multiple binding sites for CTCF, a methylation-sensitive factor that mediates chromatin insulator activity (Bell and Felsenfeld 2000; Hark et al. 2000). Another putative CTCF-dependent insulator, which may serve as an imprinted/nonimprinted boundary, was identified between *H19* and *Mrpl23/L23mrp* (Ishihara and Sasaki 2002). We therefore scanned the chicken sequence for potential CTCF sites, from *IGF2* to *MRPL23/L23MRP*. With the consensus sequence that previously identified putative insulators (Chao et al. 2002; Ishihara and Sasaki 2002), we found no potential CTCF sites in this chicken region.

Mouse *Igf2* has three DMRs, two of which are paternally methylated: DMR1 upstream of the fetal promoter contains a GCF2-binding site and acts as a methylation-sensitive silencer (Constancia et al. 2000; Eden et al. 2001); DMR2 in the last exon contains a methylation-sensitive activator (Murrell et al. 2001). The function of the maternally methylated DMR0 at the placenta-specific promoter is currently unknown. Despite our careful examination, chicken *IGF2* did not have a sequence similar to DMR0 or DMR1. However, since the 54-bp core of DMR2 is located within the protein-coding region of the last exon (Murrell et al. 2001), we were not surprised to find that it is conserved in chickens (64.8% nucleotide identity). Among the eight differentially methylated CpGs within the core, four were conserved in chickens.

Methylation status of the region corresponding to DMR2

Since the 54-bp core of *Igf2* DMR2 was the only DMR conserved in chickens, we were interested to know the methylation status of the chicken sequence. By bisulfite methylation analysis, we found that the overall methylation level at seven CpG sites (among which four were conserved) of the region is 58% (65/112), 62.5% (70/112), and 87.1% (122/140) in 5-d whole embryos (where allelic expression of *IGF2* was studied) (Fig. 7A), 7-d whole embryos and peripheral blood, respectively (Fig. 8). Although we were not able to distinguish between the parental alleles because of the lack of SNPs, most of the sites were rather uniformly methylated or unmethylated (Fig. 8), suggesting that there is no allelic bias in methylation.

Other local regulatory elements

At about 40 kb 3′ to mouse *Igf2*, there is a conserved segment called A6A4, which contains two DNaseI-hypersensitive sites

(Koide et al. 1994). A targeted deletion of this segment has led to biallelic *Igf2* expression in the brain and a relaxation of *Igf2* imprinting in the skeletal muscle, suggesting a tissue-specific silencer activity (Ainscough et al. 2000; Jones et al. 2001). Also, two DNA segments, termed *H19* upstream conserved regions (HUCs), have been shown to act as enhancers in a range of mesodermal tissues (Drewell et al. 2002). However, we were not able to identify sequences similar to these in the chicken sequence between *IGF2* and *MRPL23/L23MRP*.

We previously identified 10 conserved noncoding segments (CS1–10), which are located between *H19* and *Mrpl23/L23mrp* (Ishihara et al. 2000). It was shown by transgenic assays that seven of them possess tissue-specific enhancer activities (Ishihara et al. 2000). In fact, two were identical with the previously reported endoderm-specific enhancers shared by *Igf2* and *H19* (Leighton et al. 1995b). Although we attempted to identify sequences similar to the enhancers in the chicken *IGF2–MRPL23/L23MRP* region, no significant homology was detected. The tissue-specific enhancers in chickens, if present, may be much diverged from those in the mammalian species.

Conserved noncoding sequences

Having established that most regulatory elements identified in mammals are not conserved in chickens, we asked whether there are any noncoding homologies (excluding the promoters) between mice and chickens. A PipMaker analysis revealed several conserved noncoding sequences (Fig. 3). Table 4 summarizes those that showed highest sequence identities (>50%) over 50 bp.

There were three conserved noncoding sequences in the intergenic region between *ASCL2/CASH4* and *TH* (Fig. 3; Table 4). The sequences were 51 bp to 179 bp in size, and their identity to the corresponding mouse sequence ranged from 66.7% to 75.0%. The findings suggest that *ASCL2/CASH4*, *TH*, or both are regulated by evolutionarily conserved distant control elements. Conserved noncoding sequences were also identified in introns (Fig. 3; Table 4). *MRPL23/L23MRP* had a conserved 175-bp sequence in intron 2, which showed 58.0% identity with the corresponding 182-bp mouse sequence. *TNNT3/TNT* had two conserved intronic sequences: a 64-bp sequence in intron 2 was 65.6% identical with the corresponding mouse sequence; a 90-bp sequence in intron 5 was 60.0% identical with its mouse homolog. *LSP1* had two conserved sequences in intron 10: a 327-bp and a 154-bp sequence, which, respectively, showed 51.6% and 64.6% identity with their mouse homolog. All these intronic similarities are most probably involved in the regulation of the relevant genes.

Discussion

We have determined the sequence of a 0.5-Mb chicken region orthologous to the imprinted *Ascl2/Mash2–Igf2–H19* cluster on mouse distal Chromosome 7 and on human Chromosome 11 (Reik and Maher 1997). The region was previously mapped to chicken Chromosome 5 by fluorescent in situ hybridization (Yokomine et al. 2001). The determined sequence contained eight previously reported genes/transcripts (*ASCL2/CASH4*, *TH*, *INS*, *IGF2*, *CIGF*, *TNNT3/TNT*, *TNI*, and *CTSD*) and four newly identified genes (*CD81/TAPA1*, *PHEMX/TSSC6*, *MRPL23/L23MRP*, and *LSP1*). Comparisons between the chicken and mouse sequences revealed that most of the genes are conserved between the two species, with notable exceptions of *H19* and *Nctc1*. The conserved

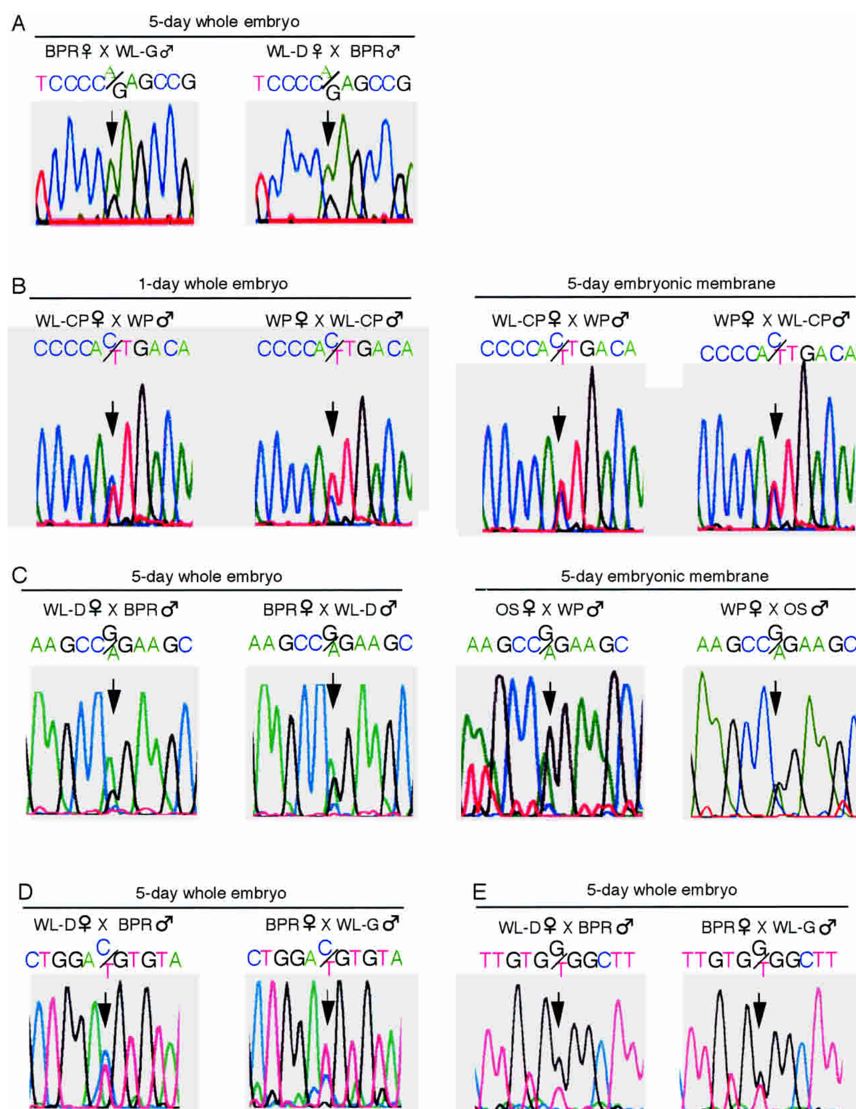

Figure 7. Imprinting status of chicken *IGF2* (*A*), *ASCL2/CASH4* (*B*), *INS* (*C*), and the control genes *PHEMX/TSSC6* (*D*) and *TH* (*E*). Biallelic expression was revealed in informative embryos and embryonic membranes including the yolk sac by RT-PCR and sequencing. The SNPs used to distinguish the parental alleles were 1242A/G for *IGF2* (Yokomine et al. 2001), 29365A/G for *ASCL2/CASH4*, 106015G/A for *INS*, 11747A/G for *PHEMX/TSSC6*, and 86744G/T for *TH*. In *B* and *D*, the data for the other strand are shown. Three additional SNPs confirmed the biallelic expression of *ASCL2/CASH4* (data not shown).

Table 2. CpG islands

Gene	Location	Length (bp)
CD81/TAPA1	Intron 1	206
ASCL2/CASH4	Upstream	304
	Upstream	773
	Upstream ~ exon 1	551
IGF2	Upstream	223
	Upstream	906
	Upstream	321
	Intron 1	252
	Exon 3	363
MRPL23/L23MRP	Upstream ~ exon 1	234
LSP1	Intron 1	491
	Intron 1	312
CTSD	Upstream ~ exon 1	442

genes maintain the same gene order, exon–intron structures, and transcriptional polarities.

Among the conserved genes, *ASCL2/CASH4* was of special interest since its mammalian ortholog *Ascl2/Mash2* is only expressed in diploid trophoblast cells and required for development of the placenta (Guillemot et al. 1994). Chicken *ASCL2/CASH4* is most probably involved in neural development (Henrique et al. 1997), and thus the gene provides a good example of functional diversification during vertebrate evolution. The role in neural development is not unexpected because the members of this basic helix–loop–helix transcription factor family are often involved in neural development in many species.

Having identified the genes/transcripts of the chicken region, we examined whether the orthologs of the mammalian imprinted genes show imprinted expression patterns. We found that chicken *ASCL2/CASH4* and *INS* are not imprinted in developing embryos or in the embryonic membranes, including the yolk sac. Together with the previous data that chicken *IGF2* and *MPR1* are not imprinted (O'Neill et al. 2000; Nolan et al. 2001; Yokomine et al. 2001), our findings support the idea that genomic imprinting does not exist in chickens. This is consistent with the conflict theory (Moore and Haig 1991), or the kinship theory (Wilkins and Haig 2003), which states that imprinting evolved as a result of conflicting interests between the paternal and maternal genes over allocation of maternal resources to the offspring. The theory predicts that imprinting would be more favored by viviparous species than oviparous species.

Thus, the sequence reported here for the first time enabled us to compare in detail a mammalian imprinted region with its orthologous region from a nonimprinted vertebrate species. The presence of the same gene cluster in the imprinted and nonimprinted species suggests that the imprinting of the genes in this cluster evolved on a domain basis, but not on a gene-by-gene basis. Several sequence elements involved in imprinting and long-distance regulation of the cluster have been identified in mice and humans. We found that such elements of the region, including the *H19* IC, DMRs, and HUCs, are not conserved in chickens. The only exception was the 54-bp core of *Igf2* DMR2, which is a part of the protein-coding region, but it did not show allele-specific differential methylation in chickens. Thus, it is likely that the IC and other local regulatory elements involved in the imprinted expression of the genes evolved after the diversification of mammals and birds.

In this connection, it is interesting that not only the *H19* IC but also *H19* itself was not found in the chicken sequence. Mammalian *H19* encodes no protein, and a germ-line deletion of mouse *H19* and its IC showed that their sole function may be to imprint *Igf2* and *Ins2* (Leighton et al. 1995a). It is therefore tempting to speculate that *H19* and the linked IC were acquired by horizontal gene transfer in an ancestral species of mammals and caused imprinting. A possible link between parasitic DNA and imprinting was discussed previously (Barlow 1993; Yoder et al. 1997). Alternatively, the common ancestor of mammals and birds might have had this gene. Then this nonessential gene was lost in birds while it acquired a new function (imprinting) in mammals.

It is also noteworthy that chickens lack the large cluster of tandem repeats and retroelements, which we recently identified in the mouse genome between *Th* and *Ins2* (Shirohzu et al. 2004). Approximately 46% of the 210-kb mouse region consists of retroelements such as LINE-1 and IAP with 30% of the remaining being tandem repeats and, despite the heterochromatin-like se-

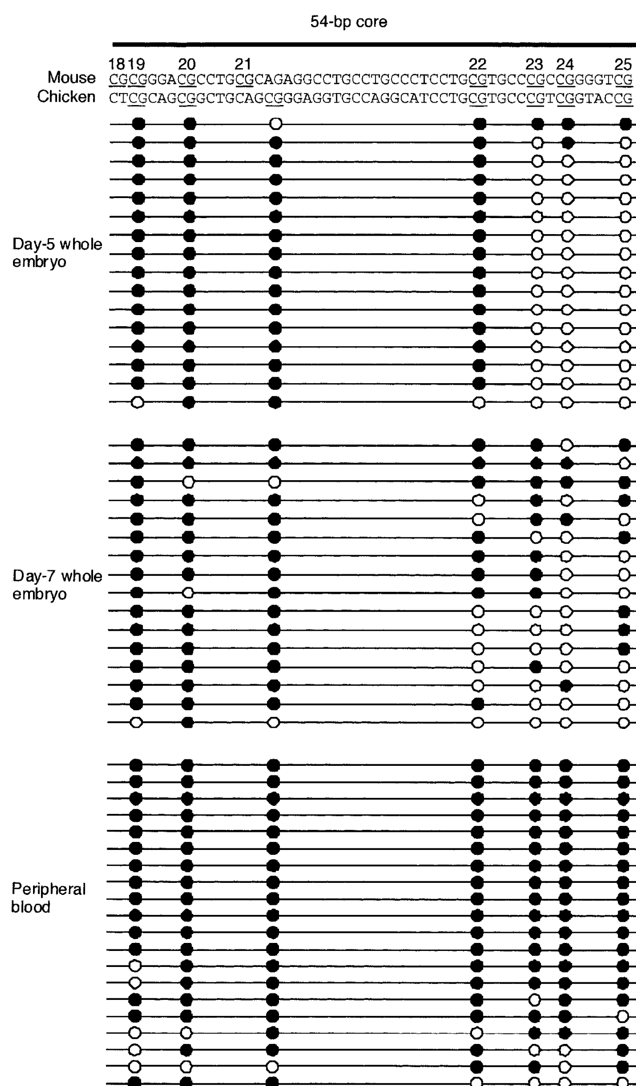

Figure 8. Bisulfite methylation analysis of the chicken *IGF2* region homologous to the 54-bp core of the mouse DMR2. CpG dinucleotides are underlined. The mouse CpG sites are numbered according to Murrell et al. (2001). Filled circles indicate methylated sites and open circles unmethylated sites. There is no indication for allelic difference in DNA methylation.

Table 3. Tandem repeats

		Position	Unit size (bp)	Consensus sequence	Copy number
Chicken	cTR1	26,747–27,120[a]	16	TGGCCATGGGGTTGAG	26
	cTR2	29,282–29,776[a]	6	TGGGGT	29
	cTR3	139,920–140,195[a]	18	AGC(G/A)TG(G/A)TGGCCTCCATC	14

[a]The nucleotide position numbers are according to AP003796.

quence composition, this region shows asynchronous replication between the parental chromosomes. Whether this repeat-rich region has any biological function remains to be tested, but the presence of a similar retroelement-rich region in humans (between *ASCL2/HASH2* and *TH*) and the lack of its equivalence in chickens suggest a functional correlation.

Recently, Walter and Paulsen (2003) found that many imprinted genes, including those in the BWS cluster, have imprinted as well as nonimprinted paralogs, which are often linked to the other imprinted clusters. This finding has led them to propose that duplications as well as translocations and transpositions dispersed the imprinted genes and clusters throughout the genome. Since we showed that the same gene cluster is conserved in mammals and chickens, such dynamic events that affect the gene arrangement must have occurred prior to the emergence of imprinting. It is likely that many regulatory elements, noncoding RNA genes, and retroelements were then brought into the cluster during early mammalian evolution and eventually caused imprinting in a common ancestor of eutherians and marsupials.

Methods

Isolation and sequencing of BAC clones

Chicken BAC clones 26D12 and 192C9 were obtained from a White Leghorn library (Hori et al. 2000) with a PCR-amplified probe from the last coding exon of chicken *IGF2* (primers: 5′-GAGAGCTTCCAGAAGCCATCTC-3′ and 5′-GCCCAACTGTCCC TTCGTAAGT-3′) (Yokomine et al. 2001) by colony hybridization. Then a 192C9-end STS (primers: 5′-CATGAGAAGTGACTTTCTG AAGCC-3′ and 5′-CCTGTCCCTGTGTTGCAGATGAG-3′) was used as a probe to isolate 27B1 and 161D9. Chicken STS primers used to assess the STS content of the BAC clones were *TH*, 5′-AG AGGACTGGCTTCCAGCTCCG-3′ and 5′-CTGGGAGAACTGG GCAAACGTCT-3′; *INS*, 5′-GGCTCTCTGGATCCGATCAC-3′ and 5′-GGCTGCTCGACATCCCGTCG-3′; *TNNT3/TNT*, 5′-GAGG AAAAGGCACGGAGAGAGG-3′ and 5′-CTTTGCCAGATAGCTG CTGTATGA-3′. A combined shotgun (Fleischmann et al. 1995) and nested-deletion (Hattori et al. 1997) strategy was adopted to sequence the BAC inserts as described (Hattori et al. 2000). Sequence data were assembled by Phred-Phrap and Sequencher software (Gene Codes).

Sequence data analysis

Database homology search and gene predictions were performed with BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) and GENSCAN (http://genes.mit.edu/GENSCAN.html), respectively. CpG islands were predicted by using cpgplot (http://bioweb.pasteur.fr/seqanal/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.

RT-PCR

For tissue distribution analysis, alternative splicing analysis, and allelic expression analysis, reverse transcription of total RNA from whole chick embryos and adult tissues was performed using M-MLV reverse transcriptase (GIBCO BRL) according to the manufacturer's protocol. PCR was carried out under the following conditions: 30 cycles of 94°C for 30 sec, appropriate annealing temperature for 30 sec, and 72°C for 30 sec, with a final extension at 72°C for 5 min. The annealing temperature was 65°C for *CD81/TAPA1*, 70°C for *PHEMX/TSSC6*, 68°C for *TH*, 65°C for *MRPL23/L23MRP* and 60°C for *LSP1*. The primers used were *CD81/TAPA1*, 5′-TTGGCCTCAGACCGGGAGCTG-3′ and 5′-GATGAAGTTGAAGACGAAGAGCAG-3′; *PHEMX/TSSC6* exons 2/7, 5′-TTTCTACGCCGGCGTCTTCCTCA-3′ and 5′-CTTCCT CGCAAGGTGTATTTGCCT-3′; *PHEMX/TSSC6* exons 3/6, 5′-TTT GCTCTGGCCTTCTGTGGGATG-3′ and 5′-CGTGGAGACAAA GCCCATGTGCTT-3′; *TH*, 5′-TGTGTCTGAGAGCTTCAGTGA TGC-3′ and 5′-GAAGCTGGCTTTCAGTAAAGCAGG-3′; *MRPL23/L23MRP*, 5′-ACCCCTTGTACCAGCTGGGTGG-3′ and 5′-AACT GCACGGTGTCCTCAGGCT-3′; *LSP1*, 5′-ATTGCTCCAGTCTGT CATCTATC-3′ and 5′-GCAGAGGAGGGCTTCAATGGCA-3′. *ASCL2/CASH4* and *INS* cDNAs were amplified by nested PCR. First PCR was carried out under the following condition: 10 cycles of 94°C for 30 sec, appropriate annealing temperature for 30 sec, and 72°C for 30 sec, with a final extension at 72°C for 5 min. The annealing temperature was 68°C for *ASCL2/CASH4* and 65°C for *INS*. The primers used were *ASCL2/CASH4*, 5′-GAGGAGCAGGAACTGCTGGATTTC-3′ and 5′-GCTGTGGAA GACCATAGGAATCGA-3′; *INS*, 5′-AACCAGCTATGCAGCTGCC AACC-3′ and 5′-GAGTAAGTGTATGTCTGTGCCCCGC-3′. Using an aliquot of the products as a template, a second PCR was done for 30 cycles with the same parameters. The primers used were *ASCL2/CASH4*, 5′-ACCAGCTGGCTTGGGAGCTACTGA-3′ and 5′-CCAATGCCTTTGACAACCTGTTGG-3′; *INS*, 5′-GGAGAG CGTGGCTTCTTCTACTC-3′ and 5′-GAGTAAGTGTATGTCTGTG CCCGC-3′.

Table 4. Conserved noncoding sequences

Gene	Location	Size (bp)	Size in mouse (bp)	Identity (%)
ASCL2/CASH4	Downstream 10 kb	115	119	68.3
	Downstream 15 kb	51	52	75.0
	Downstream 18 kb	179	167	66.7
MRPL23/L23MRP	Intron 2	175	182	58.0
TNNT3/TNT	Intron 2	64	64	65.6
	Intron 5	90	87	60.0
LSP1	Intron 10	327	303	51.6
	Intron 10	154	161	64.6

Identification of SNPs and allelic expression analysis

To identify SNPs within the transcribed regions, PCR primers were designed to amplify the exons of each gene. The primers were *PHEMX/TSSC6* exons 3 and 3′, 5′-ACCTCTCAAGAGTCGAGCAGCTC-3′ and 5′-TGCTCTAGCTTTAACCAGGCTTGC-3′; *PHEMX/TSSC6* exon 4, 5′-TGGCAAAAGGCTGGTTCTGGAGG-3′ and 5′-TGTCACAGGCACTTCTGTTTCTGTA-3′; *PHEMX/TSSC6* exon 5, 5′-TGTTGTCACACAACCTAAACACGT-3′ and 5′-ATAGCATGGTAGGCACCTGCTTAG-3′; *PHEMX/TSSC6* exon 6, 5′-TTCCATGGTGGAGGCCATCACCA-3′ and 5′-CATCCTCTGCAGACTGGAACTCAT-3′; *ASCL2/CASH4* exon 1, 5′-ACCAGCTGGCTTGGGAGCTACTGA-3′ and 5′-GCTGTGGAAGACCATAGGAATCGA-3′; *TH* exon 13, 5′-ACTACGCAGCACATATCAAGAGGC-3′ and 5′-GAAGCTGGCTTTCAGTAAAGCAGG-3′; *INS* exon 1, 5′-TCACGTCAAAGGAGCTGAGGGAC-3′ and 5′-GGACATTCCTTGTGTCACCATCAAA-3′; *INS* exon 2, 5′-CAGCTCTTCACTTACACACCTGGT-3′ and 5′-GTGGTGTCCCCTCCACAAGAAAC-3′. PCR was done using genomic DNA as a template under the following condition: 30 cycles of 94°C for 30 sec, 65°C for 30 sec, and 72°C for 30 sec, with a final extension at 72°C for 5 min. PCR products were purified using Microcon TM-100 (Millipore) and sequenced using BigDye Terminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems) and the ABI 377 DNA Sequencer.

Following appropriate crosses, total RNA was obtained from F_1 embryos or the tissues from F_1 individuals heterozygous for the SNP. Allelic expression status of *PHEMX/TSSC6* was examined by sequencing the RT-PCR products amplified with the primer set "exons 3/6." Since this primer set amplifies multiple alternative splicing products, the RT-PCR product from the most abundant isoform, isoform 5, was purified and sequenced. Allelic expression status of *ASCL2/CASH4*, *TH*, and *INS* was examined by sequencing the RT-PCR products amplified with the primers described in the "RT-PCR" section.

Bisulfite methylation analysis

For analysis, 1 µg of DNA isolated from peripheral blood was subjected to bisulfite methylation analysis (Frommer et al. 1992). The bisulfite treatment was carried out with EZ DNA Methylation Kit (Zymo Research). Semi-nested PCR was preformed to amplify the chicken region homologous to the DMR2 of mouse *Igf2*. The primers were bisIGF2-P1, 5′-GTGTTGATATTGTGTTGTTTTTTTTTT-3′; bisIGF2-P3 (outer), 5′-CCACCCCTCCTTACTTATATCATTT-3′; bisIGF2-P2 (inner), 5′-TAACTTCCTCAACTACTTACAACCC-3′. PCR was carried out under the following conditions: first round of amplification: 5 cycles of 95°C for 1 min, 50°C for 2 min, and 72°C for 3 min; and 25 cycles of 94°C for 1 min, 50°C for 2 min, and 72°C for 2 min, with a final extension at 72°C for 7 min. Second round of amplification: 25 cycles of 94°C for 30 sec, 65°C, and then reduced by 0.5°C for each later cycle for 30 sec; and 72°C for 30 sec and 10 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 TOPO TA Cloning System (Invitrogen), and sequenced using BigDye Terminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems) and ABI 377 DNA Sequencer.

Acknowledgments

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 NIG Collaboration Research Program. We thank the members of Division of Human Genetics, Department of Integrated Genetics, National Institute of Genetics and Human Genome Research Group, Genomic Sciences Center, RIKEN, for technical support and discussions.

References

Ainscough, J.F-X., John, R.M., Barton, S.C., and Surani, M.A. 2000. A skeletal muscle-specific mouse *Igf2* repressor lies 40 kb downstream of the gene. *Development* **127:** 3923–3930.

Andria, M.L., Hsieh, C.L., Oren, R., Francke, U., and Levy, S. 1991. Genomic organization and chromosomal localization of the *TAPA-1* gene. *J. Immunol.* **147:** 1030–1036.

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

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

Benson, G. 1999. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **27:** 573–580.

Brannan, C.I., Dees, E.C., Ingram, R.S., and Tilghman, S.M. 1990. The product of the *H19* gene may function as an RNA. *Mol. Cell. Biol.* **10:** 28–36.

Brown, W.R.A., Hubbard, S.J., Tickle, C., and Wilson, S.A. 2003. The chicken as a model for large-scale analysis of vertebrate gene function. *Nat. Rev.* **4:** 87–97.

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.

Cattanach, B.N. and Beechey, C.V. 1997. Genomic imprinting in the mouse: Possible final analysis. In *Genomic imprinting* (eds. W. Reik and A. Surani), pp. 118–145. IRL Press at Oxford University Press, Oxford, UK.

Chao, W., Huyuh, K.D., Spencer, R.J., Davidow, L.S., and Lee, J.T. 2002. CTCF, a candidate *trans*-acting factor for X-inactivation choice. *Science* **295:** 345–347.

Constancia, M., Dean, W., Lopes, S., Moore, T., Kelsey, G., and Reik, W. 2000. Deletion of a silencer element in *Igf2* results in loss of imprinting independent of *H19*. *Nat. Genet.* **26:** 203–206.

Drewell, R.A., Arney, K.L., Arima, T., Barton, S.C., Brenton, J.D., and Surani, M.A. 2002. Novel conserved elements upstream of the *H19* gene are transcribed and act as mesodermal enhancers. *Development* **129:** 1205–1213.

Eden, S., Constancia, M., Hashimshony, T., Dean, W., Goldstein, B., Johnson, A.C., Keshet, I., Reik, W., and Cedar, H. 2001. An upstream repressor element plays a role in *Igf2* imprinting. *EMBO J.* **20:** 3518–3525.

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

Ferguson-Smith, A.C. and Surani, M.A. 2001. Imprinting and the epigenetic asymmetry between parental genomes. *Science* **293:** 1086–1089.

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 *KvDMR1*. *Nat. Genet.* **32:** 426–431.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M., et al. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269:** 496–512.

Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L., and Paul, C.L. 1992. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci.* **89:** 1827–1831.

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

Giddings, S.J., King, C.D., Harman, K.W., Flood, J.F., and Carnaghi, L.R. 1994. Allele specific inactivation of insulin 1 and 2, in the yolk sac, indicates imprinting. *Nat. Genet.* **6:** 310–313.

Guillemot, F., Nagy, A., Auerbach, A., Rossant, J., and Joyner, A.L. 1994. Essential role of *Mash-2* in extraembryonic development. *Nature* **371:** 333–336.

Guillemot, F., Caspary, T., Tilghman, S.M., Copeland, N.G., Gilbert, D.J., Jenkins, N.A., Anderson, D.J., Joyner, A.L., Rossant, J., and Nagy, A. 1995. Genomic imprinting of *Mash2*, a mouse gene required for trophoblast development. *Nat. Genet.* **9:** 235–242.

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.

Hattori, M., Tsukahara, F., Furuhata, Y., Tanahashi, H., Hirose, M., Saito,

M., Tsukuni, S., and Sakaki, Y. 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.
Hattori, M., Fujiyama, A., Taylor, T.D., Watanabe, H., Yada, T., Park, H.S., Toyoda, A., Ishii, K., Totoki, Y., Choi, D.K., et al. 2000. The DNA sequence of human chromosome 21. *Nature* **405:** 311–319.
Henrique, D., Tyler, D., Kintner, C., Heath, J.K., Lewis, J.H., Ish-Horowicz, D., and Storey, K.G. 1997. *cash4*, a novel *achaete–scute* homolog induced by Hensen's node during generation of the posterior nervous system. *Genes & Dev.* **11:** 603–615.
Hori, T., Asakawa, S., Itoh, Y., Shimizu, N., and Mizuno, S. 2000. *Wpkci*, encoding an altered form of PKCI, is conserved widely on the avian W chromosome and expressed in early female embryos: Implication of its role in female sex determination. *Mol. Biol. Cell* **11:** 3645–3660.
Horike, S., Mitsuya, K., Meguro, M., Kotobuki, N., Kashiwagi, A., Notsu, T., Schulz, T.C. Shirayoshi, Y., and Oshimura, M. 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.
Ishihara, K. and Sasaki, H. 2002. An evolutionarily conserved putative insulator element near the 3′ boundary of the imprinted *Igf2/H19* domain. *Hum. Mol. Genet.* **11:** 1627–1636.
Ishihara, K., Kato, R., Furuumi, H., Zubair, M., and Sasaki, H. 1998. Sequence of a 42-kb mouse region containing the imprinted *H19* locus: Identification of a novel muscle-specific transcription unit showing biallelic expression. *Mamm. Genome* **9:** 775–777.
Ishihara, K., Hatano, N., Furuumi, H., Kato, R., Iwaki, T., Miura, K., Jinno, Y., and Sasaki, H. 2000. Comparative genomic sequencing identifies novel tissue-specific enhancers and sequence elements for methylation-sensitive factors implicated in *Igf2/H19* imprinting. *Genome Res.* **10:** 664–671.
Jones, B.K., Levorse, J., and Tilghman, S.M. 2001. Deletion of a nuclease-sensitive region between the *Igf2* and *H19* genes leads to *Igf2* misregulation and increased adiposity. *Hum. Mol. Genet.* **10:** 807–814.
Jongstra, J., Tidmarsh, G.F., Jongstra-Bilen, J., and Davis, M.M. 1988. A new lymphocyte-specific gene which encodes a putative Ca^{2+}-binding protein is not expressed in transformed T-lymphocyte lines. *J. Immunol.* **141:** 3999–4004.
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.
Killian, J.K., Byrd, J.C., Jirtle, J.V., Munday, B.L., Stoskopf, M.K., MacDonald, R.G., and Jirtle, R.L. 2000. *M6P/IGF2R* imprinting evolution in mammals. *Mol. Cell* **5:** 707–716.
Koide, T., Ainscough, J., Wijerde, M., and Surani, M.A. 1994. Comparative analysis of *Igf2/H19* imprinted domain: Identification of a highly conserved intergenic DNase I hypersensitive region. *Genomics* **24:** 1–8.
Larsen, F., Gundersen, G., Lopez, R., and Prydz, H. 1992. CpG islands as gene markers in the human genome. *Genomics* **13:** 1095–1107.
Leighton, P.A., Ingram, R.S., Eggenschwiler, J., Efstratiadis, A., and Tilghman, S.M. 1995a. Disruption of imprinting caused by deletion of the *H19* gene region in mice. *Nature* **375:** 34–39.
Leighton, P.A., Saam, J.R., Ingram, R.S., Stewart, C.L., and Tilghman, S.M. 1995b. An enhancer deletion affects both *H19* and *Igf2* expression. *Genes & Dev.* **9:** 2079–2089.
Li, E. 2002. Chromatin modification and epigenetic reprogramming in mammalian development. *Nat. Rev. Genet.* **3:** 662–673.
Misener, V.L., Wielowieyski, A., Brennan, L.A., Beebakhee, G., and Jongstra, J. 1998. The mouse *Lsp1* and *Tnnt3* genes are 4.3 kb apart on distal mouse chromosome 7. *Mamm. Genome* **9:** 846–848.
Miyamoto, T., Hasuike, S., Jinno, Y., Soejima, H., Yun, K., Miura, K., Ishikawa, M., and Niikawa, N. 2002. The human *ASCL2* gene escaping genomic imprinting and its expression pattern. *J. Assist. Reprod. Genet.* **19:** 240–244.
Moore, T. and Haig, D. 1991. Genomic imprinting in mammalian development: A parental tug-of-war. *Trends Genet.* **7:** 45–49.
Moore, G.E., Abu-Amero, S.N., Bell, G., Wakeling, E.L., Kingsnorth, A., Stanier, P., Jauniaux, E., and Bennett, S.T. 2001 Evidence that insulin is imprinted in the human yolk sac. *Diabetes* **50:** 199–203.
Murrell, A., Heeson, S., Bowden, L., Constancia, M., Dean, W., Kelsey, G., and Reik, W. 2001. An intragenic methylated region in the imprinted *Igf2* gene augments transcription. *EMBO Rep.* **2:** 1101–1106.
Neumann, B., Kubicka, P., and Barlow, D.P. 1995. Characteristics of imprinted genes. *Nat. Genet.* **9:** 12–13.
Nolan, C.M., Killian, J.K., Petitte, J.N., and Jirtle, R.L. 2001. Imprint status of *M6P/IGF2R* and *IGF2* in chickens. *Dev. Genes Evol.* **211:** 179–183.
Okamura, K., Hagiwara-Takeuchi, Y., Li, T., Vu, T.H., Hirai, M., Hattori, M., Sakaki, Y., Hoffman, A.R., and Ito, T. 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.
O'Neill, M.J., Ingram, R.S., Vrana, P.B., and Tilghman S.M. 2000. Allelic expression of *IGF2* in marsupials and birds. *Dev. Genes Evol.* **210:** 18–20.
Onyango, P., Miller, W., Lehoczky, J., Leung, C.T., Birren, B., Wheelan, S., Dewar, K., and Feinberg, A.P. 2000. Sequence and comparative analysis of the mouse 1-megabase region orthologous to the human 11p15 imprinted domain. *Genome Res.* **10:** 1697–1710.
Paulsen, M., El-Maarri, O., Engemann, S., Strodicke, M., Franck, O., Davies, K., Reinhardt, R., Reik, W., and Walter, J. 2000. Sequence conservation and variability of imprinting in the Beckwith-Wiedemann syndrome gene cluster in human and mouse. *Hum. Mol. Genet.* **9:** 1829–1841.
Reik, W. and Maher, E.R. 1997. Imprinting in clusters: Lessons from Bechwith-Wiedemann syndrome. *Trends Genet.* **13:** 330–334.
Reik, W. and Walter, J. 2001. Genomic imprinting: Parental influence on the genome. *Nat. Rev. Genet.* **2:** 21–32.
Schwartz, S., Zhang, Z., Frazer, K.A., Smit, A., Riemer, C., Bouck, J., Gibbs, R., Hardison, R., and Miller, W. 2000. PipMaker—A web server for aligning two genomic DNA sequences. *Genome Res.* **10:** 577–586.
Shirohzu, H., Yokomine, T., Sato, C., Kato, R., Toyoda, A., Purbowasito, W., Suda, C., Mukai, T., Hattori, M., Okumura, K., et al. 2004. A 210-kb segment of tandem repeats and retroelements located between imprinted subdomains of mouse distal chromosome 7. *DNA Res.* **11:** 325–334.
Solter, D. 1988. Differential imprinting and expression of maternal and paternal genomes. *Annu. Rev. Genet.* **22:** 127–146.
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* **Suppl.** 89–98.
Taylor, E.R., Seleiro, E.A.P., and Brickell, P.M. 1991. Identification of antisense transcripts of the chicken insulin-like growth factor-II gene. *J. Mol. Endocrinol.* **7:** 145–154.
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.
Tilghman, S.M. 1999. The sins of the fathers and mothers: Genomic imprinting in mammalian development. *Cell* **96:** 185–193.
Tsang, P., Gilles, F., Yuan, L., Kuo, Y.-H., Lupu, F., Samara, G., Moosiksuwan, J., Goye, A., Zelenetz, A.D., Selleri, L., et al. 1995. A novel L23-related gene 40-kb downstream of the imprinted *H19* gene is biallelically expressed in mid-fetal and adult human tissues. *Hum. Mol. Genet.* **4:** 1499–1507.
Walter, J. and Paulsen, M. 2003. The potential role of gene duplications in the evolution of imprinting mechanisms. *Hum. Mol. Genet.* **12:** R215–R220.
Wilkins, J.F. and Haig, D. 2003. What good is genomic imprinting: The function of parent-specific gene expression. *Nat. Rev. Genet.* **4:** 359–368.
Yoder, J.A., Walsh, C.P., and Bestor, T.H. 1997. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* **13:** 335–340.
Yokomine, T., Kuroiwa, A., Tanaka, K., Tsudzuki, Y., Matsuda, Y., and Sasaki H. 2001 Sequence polymorphisms, allelic expression status and chromosome locations of chicken *IGF2* and *MPR1* genes. *Cytogenet. Cell Genet.* **93:** 109–113.
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.
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.

Web site references

http://bioweb.pasteur.fr/seqanal/interfaces/cpgplot.html; CpG plot.
http://c3.biomath.mssm.edu/trf.html; Tandem Repeats Finder.
http://genes.mit.edu/GENSCAN.html; GENSCAN.
http://nog.cse.psu.edu/pipmaker/; PipMaker.
http://repeatmasker.genome.washington.edu/cgi-bin/RepeatMasker; RepeatMasker.
http://swallow.gsf.de/dt40est.html; BursaEst Database.
http://www.ncbi.nlm.nih.gov/BLAST/; BLAST.

Received March 22, 2004; accepted in revised form August 14, 2004.