SHORT COMMUNICATION

Exon Sharing of a Novel Human Zinc-Finger Gene, *ZIM2*, and Paternally Expressed Gene 3 (*PEG3*)

Joomyeong Kim,*,†,[1] Anne Bergmann,* and Lisa Stubbs*,†

*Human Genome Center, Biology and Biotechnology Research Program, L-452, Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, California 94551; and †DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, California 94598

Received October 14, 1999; accepted December 20, 1999

We have identified a novel human gene, *ZIM2* (zinc-finger gene 2 from imprinted domain), located 25 kb downstream of *PEG3* (paternally expressed gene 3). *ZIM2* produces two different-size transcripts, 2.5 and 9.0 kb in length, with highest levels of expression in adult testis and modest levels in fetal kidney and brain. The 2.5-kb transcript of *ZIM2* consists of 11 exons and encodes a Kruppel-type (C2H2) zinc-finger protein with a conserved Kruppel-associated box (KRAB) domain. Rapid amplification of cDNA ends and cDNA sequencing studies showed that *ZIM2* and *PEG3* transcripts share identical 5′-ends, composed of 7 small exons. Alternative splicing events connect these 7 exons either with the remaining 2 exons of *PEG3* or with the remaining 4 exons of *ZIM2*. Interestingly, the third among the 7 shared exons exhibits sequence similarity to leucine-rich domains that are found at the N-terminal region of a subset of KRAB-containing zinc-finger genes. Sequencing of the 5′-termini of both transcripts indicates that *ZIM2* and *PEG3* share identical transcription start sites and may also share upstream regulatory elements, although the two genes show distinct patterns of tissue-specific expression.

Peg3 (paternally expressed gene 3) is the first imprinted gene identified from the proximal region of mouse chromosome 7 (7). *Peg3* encodes a Cys2-His2 zinc-finger protein, most of which are thought to act as transcription factors (13, 14). Recently, the *in vivo* mutational study of *Peg3* has demonstrated a role for this gene in determining the maternal caring behavior of female adults (8). The human homolog of *Peg3* has been previously mapped to a genomic region in chromosome 19q13.4, where a large number of Kruppel-type (C2H2) zinc-finger genes (ZNFs) are clustered (5). Since imprinted genes tend to be clustered in chromosomal regions (1, 9, 12), we have examined the structure, expression, and imprinting status of ZNF genes that are located close to human and mouse *PEG3*. Using this approach, a maternally expressed mouse imprinted gene, *Zim1* (imprinted zinc-finger gene 1), was identified recently (6). In this study, we describe a new human zinc-finger gene, *ZIM2* (zinc-finger gene 2 from imprinted domain), located near *PEG3* and report the genomic organization and expression patterns of *ZIM2* in human tissues.

During initial studies of human *PEG3* (5), we isolated one evolutionarily conserved genomic fragment, 14378Krab, from the region downstream of *PEG3* (6). This genomic fragment encodes a potential Kruppel-associated box A domain (KRAB A), which is often found at the N-terminus of Kruppel-type (C2H2) zinc-finger proteins (2, 14). By examining the completed sequence of a human BAC clone containing *PEG3* (GenBank Accession No. AC006115), we were able to locate the KRAB fragment sequence 25 kb downstream of human *PEG3*. Careful inspection of the surrounding areas also identified a putative zinc-finger-encoding region. Subsequent BLAST analyses with this finger sequence identified a single EST entry (GenBank Accession No. AA883959) in the database, suggesting the presence of an actively expressed ZNF gene, hereafter named *ZIM2*. The EST clone, about 300 bp in length, contains only part of the finger region and 3′-UTR region of *ZIM2*. Therefore a series of RT-PCR and rapid amplification of cDNA ends (RACE) experiments were performed to obtain full-length cDNAs for *ZIM2*. First, a 1.2-kb ZIM2 cDNA was amplified from a testis cDNA template (human testis Marathon cDNA template; Clontech) by RT-PCR with two oligonucleotide primers: 14378KRAB-A (5′-TGACCTTCGAGGATGTGCTT-3′) and ZIM2-4 (5′-CAGTGATCGCACTCAACAGT-3′). Second, both 5′- and 3′-ends of *ZIM2* cDNA were further extended through the RACE technique (4) with the following two primer sets: ZIM2-4 and 14378KRAB-B (5′-CCAGGGAGACCAGTTCCGGT-3′)

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession No. AF166122.

[1] To whom correspondence should be addressed. Telephone: (925) 424-3792. Fax: (925) 422-2282. E-mail: kim16@llnl.gov.

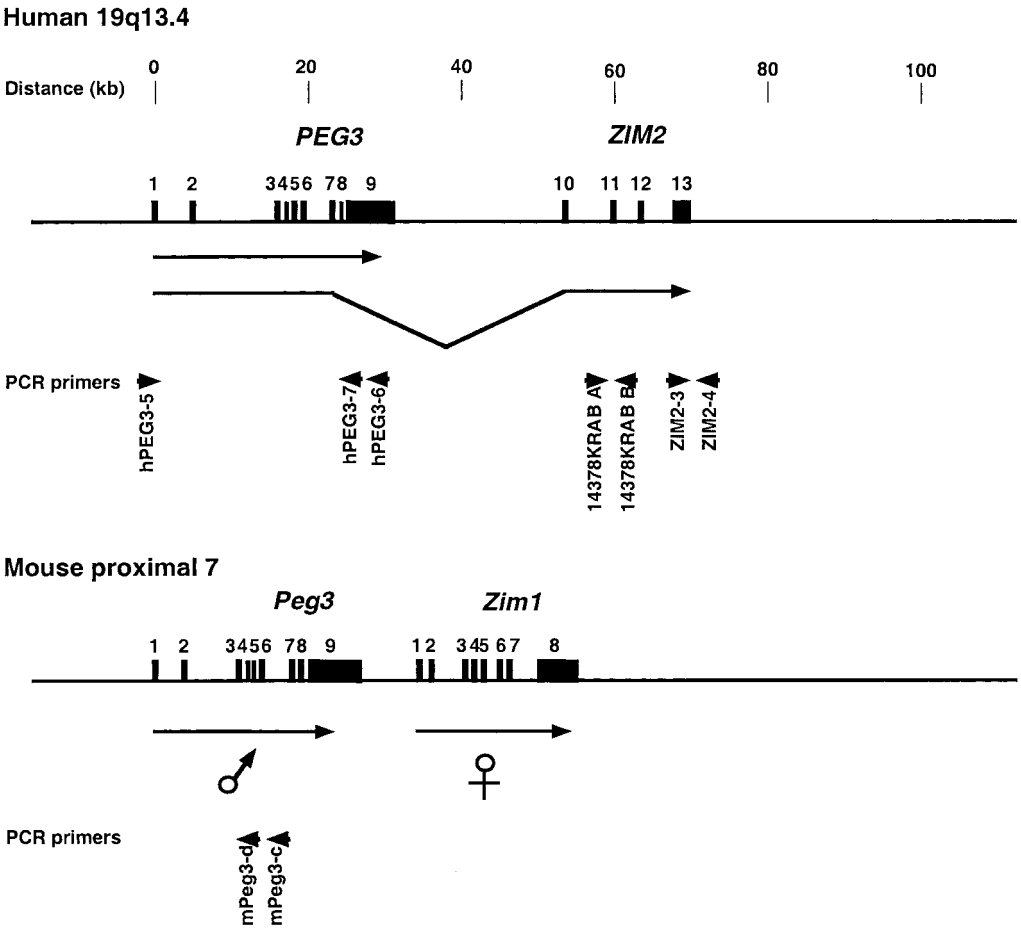

FIG. 1. Genomic organization of human *PEG3/ZIM2* and mouse *Peg3/Zim1*. The exons are depicted by black boxes. The transcriptional direction of each gene is indicated by an arrow. The transcribed parental allele of each imprinted gene is indicated by a sex symbol underneath the arrow. The positions and directions of oligonucleotide primers used for this study are also indicated by small arrowheads. The 5′-ends of human and mouse *PEG3* were determined with two separate RACE experiments using the following oligonucleotides: hPEG3-6 (5′-CACTGAACGACCTCCCACAC-3′) and hPEG3-7 (5′-ATGCATGATCTGGTGCTCAA-3′) for human *PEG3* and mPeg3-c (5′-CGAGACTCATAATCCATGGAACGTC-3′) and mPeg3-d (5′-TCTGTCCAGTCCAAAATGTGGTCTT-3′) for mouse *Peg3*.

for 5′-RACE and 14378KRAB-A and ZIM2-3 (5′-GAGAGGCCTTACCAGTGTCA-3′) for 3′-RACE (Fig. 1). Overall we were able to clone and sequence a 2.1-kb cDNA corresponding to *ZIM2* (GenBank Accession No. AF166122). Analyses of the sequence of ZIM2 cDNA revealed one open reading frame, 527 amino acids in length (Fig. 2A), composed of a central KRAB A domain and five C-terminal ZNF units. In contrast to other Kruppel-type ZNFs, the finger domain of *ZIM2* contains two degenerate finger units (one is between fingers 2 and 3 and the other is between fingers 3 and 4).

Sequence comparison of ZIM2 cDNA with the human genomic sequence, surprisingly, indicated that the first seven exons of *ZIM2* are located upstream of *PEG3*, while the remaining four exons are located downstream of this gene (Fig. 1). Because of the unusual arrangement of the exons, we tested whether the 5′-most seven exons of *ZIM2* are also part of the *PEG3* transcript by performing independent RACE experiments using two oligonucleotides, hPEG3-6 and -7, derived from a unique region of *PEG3*. These 5′-RACE experiments identified eight small exons located upstream of the reported human *PEG3* cDNA sequence, a single block of 7 kb corresponding to exon 9 in Fig. 1 (GenBank Accession No. AB006625). This result confirmed that the first seven exons of *ZIM2* are indeed identical to unidentified 5′-exons of *PEG3*. Our two separate 5′-RACE experiments with oligonucleotide primers designed from the unique region of either *PEG3* or *ZIM2* all derived an identical 5′-end, suggesting that transcription of the two genes may start at the same site. These data suggest that two genes may also share regulatory elements for transcription.

The homologous mouse region, which contains *Peg3* and *Zim1*, has also been analyzed in detail (Fig. 1). Five of the novel human *PEG3/ZIM2* exons identified in RACE experiments are homologous to cDNA sequences previously described for mouse *Peg3*. These studies reported a 5.5-kb mouse *Peg3* cDNA sequence (7; GenBank Accession No. AF038939), which corresponds to exons 3–8 and part of exon 9 (8). 5′-RACE experiments with two primers, mPeg3-c and -d, derived from the mouse *Peg3* sequence identified two more exons upstream of the known exons. Our data

A

```
1---MYQPE                                          Exon 3
   DDNNSDVTSDDDMTRNRRESSPPHSVHSFS                  Exon 4
   GDRDWDRRGRSRDMEPRDRWSHTRNPRS                    Exon 5
   MPPRDLSLPVVAKTSFEMDREDDRDSRAYESRSQ              Exon 6
   DAESYQNVVDLAEDRKPHNTIQDNMENYRKLLSL              Exon 7

        (PEG3-unique)                              Exon 8
        (PEG3-unique)                              Exon 9

   GFLAQDSVPAEKRNTEMLDNLPSAGSQ                     Exon 10
   FPDFKHLGTFLVFEELVTFEDVLVDFSPEELSS               Exon 11 (KRAB)
   LSAAQRNLYREVMLENYRNLVSL
   GHQFSKPDIISRLEEEESYAMETDSRHTVICQ                Exon 12
   GESHDDPLEPHQGNQEKL                              Exon 13 (Fingers)
   LTPITMNDPKTLTPERSYGSDEFERSSN                    |
   LSKQSKDPLGKDPQEGTAPGICTSPQSA                    |
   SQENKHNRCEFCKRTFSTQVALRRHERI                    |- Finger1
   HTGKKPYECKQCAEAFYLMPHLNRHQKT                    |- Finger2
   HSGRKTSGCNEGRKPSVQCANLCERVRI                    |
   HSQEDYFECFQCGKAFLQNVHLLQHLKA                    |- Finger3
   HEAARVLPPGLSHSKTYLIRYQRKHD                      |
   YVGERACQCCDCGRVFSRNSYLIQHYRT                    |- Finger4
   HTQERPYQCQLCGKCFGRPSYLTQHYQL                    |- Finger5
   HSQEKTVECDHC--527
```

B

Leucine-rich domain

```
PEG3/ZIM2        -50 PETRTKEEIIELLVLEQYLTIIPEKLKPWVRAKKPENCEKLVTLLENYKEMYQPE  5
A56560(Zfp38)    153 ..IH...Q.L.......F...L.RE.QT..QQHC..SA.EA.....DLEQ       202
P17028(ZNF24)    83  ...H...Q.L..V....FVA.L.KE.QT...DHH...G.EA..V..D          129
AL021997(Zfp47)77  ..MHS..Q.L.......F...L.GN.QS...EQH..SG.EV.V...            122
U62907(Zfp95)    81  ..LH...Q.L.......F...L..EFQA...EHH..SG.EA.AVI.S          127
U62908(Zfp96)    77  .DLNS..Q.L.......F...L.GE.QA..QEQN..SV.EV..V..D          123
D88827(FPM315)   72  ..M....Q.L.......F...L.QEIQSR.QELH..SG.EA...V.            117
Q07230(Zfp29)    89  ..VH...----------.M..VL.REIQA.LQEHR..SS.EA.A.V.DLTQTFR   132

mPeg3            -48 VLQQLGGWSRQAFPARGDQL....T...-LGVS.TD.....T.P..SF.-MYHQE  5
```

FIG. 2. (**A**) The amino acid sequence of *ZIM2*. The conceptual amino acid sequence of *ZIM2* starts at exon 3 and ends at exon 13. The shared region between *ZIM2* and *PEG3* is exon 1 through 7. The last two exons of *PEG3* (exon 8 and 9) are skipped in the *ZIM2* protein. The two single-letter symbols, C (cysteine) and H (histidine), in the finger domain are in boldface type. (**B**) Comparison of the LeR domain of human *PEG3/ZIM2* with other LeR representatives. The LeR domain of human *PEG3/ZIM2* (exon 3) is shown with the LeRs found in other ZNFs. The GenBank accession numbers and positions of these LeRs are indicated. The first in-frame initiation codon and the following amino acids (MYQPE) in exon 3 of human and mouse *PEG3*/*ZIM2* are in boldface type.

also indicated that exon 9 of mouse *Peg3* is approximately 8 kb in length based on the detection of a poly(A) signal 3 kb downstream of the reported 5.5-kb sequence (Kim *et al.*, unpublished results). The combined length of this large 3′ block and the upstream exons totals 9 kb, consistent with the size of the major transcript detected by human and mouse *Peg3* probes (5). The overall exon structure of *Peg3* is well conserved between human and mouse: the first two exons are located more distantly upstream, whereas exons 3 through 6 are clustered in a small region, approximately 2 kb in length, in both species. Sequence comparison of newly identified 5′-exons of human and mouse *PEG3* shows a high level of conservation, about 80% nucleotide sequence identity over the 800-bp region.

Contradictory to our initial report regarding the relative gene orientation of *Zim1* and *Peg3* transcription units (6), further analyses showed clearly that *Zim1* lies in the same transcriptional direction as *Peg3* (Fig. 1). The relative location and orientation of mouse *Zim1* and *Peg3* are very similar to those of human *ZIM2* and *PEG3*. However, comparative analyses of mouse *Zim1* and human *ZIM2* revealed that the two genes share only approximately 30% amino acid sequence identity, suggesting that they are only distantly related. The low level of sequence homology between the coding sequences of *Zim1* and *ZIM2* led us to predict that the genomic interval between *PEG3* and neighboring genes would also not be conserved in the two species. However, evidence of the shared transcriptional orientation and the close proximity of mouse *Peg3* to *Zim1* (a 7-kb distance from the poly(A) signal site of *Peg3* to the first exon of *Zim1*) prompted us to test whether *Zim1* might also share exons with *Peg3*. The results of several RT-PCR and RACE experiments indicated no exon sharing in mouse (data not shown). Therefore, our data may suggest that *Zim1* and *ZIM2* are not orthologous

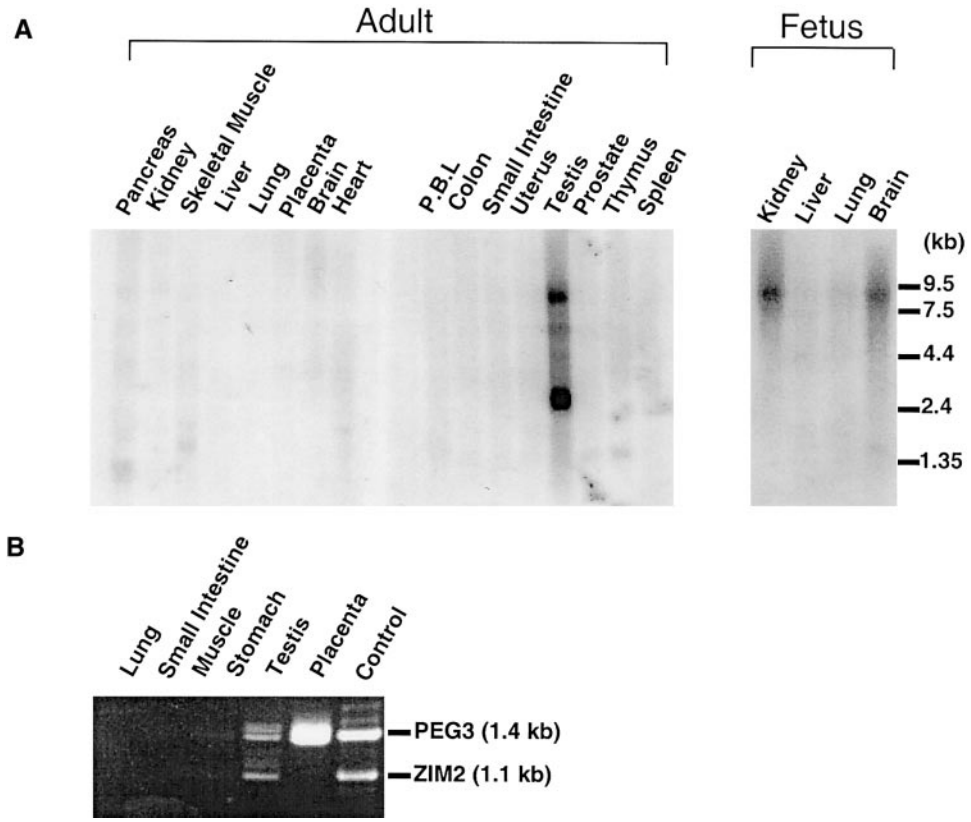

FIG. 3. *ZIM2* expression. (**A**) Northern blot analyses of *ZIM2*. Each lane contains 2 μg of poly(A)$^+$ RNA. The finger region of *ZIM2* was used as a probe. P.B.L. stands for peripheral blood leukocytes. (**B**) Competitive RT-PCR analyses of *PEG3* and *ZIM2*. Normalized cDNA templates derived from different adult tissues (human rapid-scan; OriGene Technologies) were used to measure the relative amount of *PEG3* to *ZIM2* transcripts.

genes and further, that exon sharing between *PEG3* and its nearest downstream neighbor is a unique feature of the human gene.

The open reading frames of human *PEG3* and *ZIM2* start at the first in-frame initiation codon located in exon 3; this feature also appears to be true for mouse *Peg3*. The region immediately upstream of that initiation codon, however, presumably within the 5′-UTR of *PEG3* and *ZIM2*, is a region containing an open reading frame that exhibits unusual homology with another evolutionary conserved domain, termed the leucine-rich (LeR) domain (10) (encoded by exon 3). Sequence searches with the translated amino acid sequence of exon 3 identified a large number of matched protein entries in the database, most of them being Kruppel-type ZNF genes (Fig. 2B). However, sequence analysis of RACE products provided no indication of another in-frame initiation codon upstream of exon 3. Exon 3 of mouse *Peg3* is quite diverged relative to the human exon 3 sequence; it contains several deletions and thus encodes an amino acid sequence that no longer shows significant similarity to the LeR domain (Fig. 2B). Although a functional role cannot be ruled out at present, these observations suggest that the LeR domain may represent a relic from the evolutionary history of the *PEG3* region.

To examine the expression patterns of *ZIM2*, human Northern blots containing poly(A)$^+$ RNA isolated from fetal and adult tissues were hybridized with a 1-kb cDNA corresponding to the finger region (exon 13) of *ZIM2*, a region unique to *ZIM2* (Fig. 3A). Transcripts of 2.5 and 9.0 kb were detected. The 2.5-kb transcript is expressed only in adult testis and is present at a much higher level than the 9-kb transcript in this tissue. In contrast, the 9-kb transcript is expressed at modest levels in adult testis as well as fetal kidney and brain (Fig. 3A). RACE experiments failed to yield additional *ZIM2* transcribed sequences, and the 9.0-kb transcript detected by *ZIM2* probes on Northern blots remains to be characterized. In contrast to the testis-specific expression of *ZIM2*, *PEG3* is expressed at high levels in ovary, testis, and placenta and also at modest levels in brain and heart (5). This difference was further tested with competitive RT-PCR experiments, designed to measure the relative expression levels of *PEG3* and *ZIM2* by amplifying the shared region of *PEG3* and *ZIM2*. Three different oligonucleotide primers were included in the same reaction tubes of this RT-PCR analysis: hPEG3-5 (5′-AACCACCCTTTTGAATCTGA-3′) derived from exon 1 as a 5′ primer for both *PEG3* and *ZIM2*, hPEG3-7 as a 3′primer specific for *PEG3*, and 14378KRAB-B as a 3′primer specific for *ZIM2*. The detection of *PEG3* expression in placenta, testis, and brain (data not shown) and the *ZIM2* expression in testis are consistent with the results derived from Northern blot analyses of *PEG3* and *ZIM2*. The rela-

tive amount of *PEG3* to *ZIM2* transcripts in testis appears to be almost equal based on the result of this experiment (Fig. 3B). This RT-PCR test also demonstrated independently the exon sharing of *PEG3* and *ZIM2*. According to our cDNA cloning experiments, *PEG3* and *ZIM2* might share regulatory elements; however, the two genes do show quite different expression patterns. This difference might be caused by factors other than different transcriptional regulatory elements. One possible explanation would be different rates of splicing and/or polyadenylation reactions between various tissues (3).

The predicted *ZIM2* protein contains a KRAB A and Kruppel-type (C2H2) zinc-finger domains, suggesting the functional role of *ZIM2* as a transcription factor (13, 14). 5′-Exons of *ZIM2* are identical to previously unidentified 5′-exons of human *PEG3*, demonstrating exon sharing between the two genes. A similar exon sharing has also been observed in the *Gnas* locus, where three splicing variants of *Gnas* share most of the protein-coding exons but all three are derived from different promoters (11). It is also interesting to know that *Gnas* is imprinted. Although the genomic layout of the *PEG3–ZIM2* region is very similar to that of the mouse *Peg3–Zim1* interval, the exon sharing that characterizes *PEG3* and *ZIM2* is not observed for *Peg3* and *Zim1* in mouse. These data together with the low degree of homology between *ZIM2* and *Zim1* sequences suggest that significant changes have taken place in this genomic interval since the divergence of human and mouse. One of exons shared by *PEG3* and *ZIM2* is similar to the LeR domains often found in the N-terminus of KRAB-ZNFs. Considering the close location of *PEG3* to numerous KRAB-ZNF genes and the presence of the LeR domain in one of PEG3's exons, it is possible that *PEG3* may have either been derived from or coevolved with KRAB-ZNFs.

ACKNOWLEDGMENTS

We thank Dr. Elbert Branscomb for helpful discussion and for critically reading the manuscript. This work was performed under the auspices of the U.S. Department of Energy by LLNL under Contract W-7405-ENG-48.

REFERENCES

1. Bartolomei, M. S., and Tilghman, S. M. (1997). Genomic imprinting in mammals. *Annu. Rev. Genet.* **31:** 493–525.
2. Bellefroid, E. J., Poncelet, D. A., Lecocq, P. J., Revelent, O., and Martial, J. A. (1991). The evolutionarily conserved Kruppel-associated box domain defines a subfamily of eukaryotic multifingered proteins. *Proc. Natl. Acad. Sci. USA* **88:** 3608–3612.
3. Edwards-Gilbert, G., Veraldi, K. V., and Milcarek, C. (1997). Alternative poly(A) site selection in complex transcription units: Means to an end. *Nucleic Acids Res.* **13:** 2547–2561.
4. Frohman, M. A., Dush, M. K., and Martin, G. R. (1988). Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* **85:** 8998–9002.
5. Kim, J., Ashworth, L., Branscomb, E., and Stubbs, L. (1997). The human homolog of a mouse-imprinted gene, *Peg3*, maps to a zinc finger gene-rich region of human chromosome 19q13.4. *Genome Res.* **7:** 532–540.
6. Kim, J., Lu, X., and Stubbs, L. (1999). *Zim1*, a maternally expressed mouse Kruppel-type zinc-finger gene located in proximal chromosome 7. *Hum. Mol. Genet.* **8:** 847–854.
7. Kuroiwa, Y., Kaneko-Ishino, T., Kagitani, F., Kohda, T., Li, L.-L., Tada, M., Suzuki, R., Yokoyama, M., Shiroishi, T., Wakana, S., Barton, S. C., Ishino, F., and Surani, M. A. (1996). *Peg3* imprinted gene on proximal chromosome 7 encodes for a zinc finger protein. *Nat. Genet.* **12:** 186–190.
8. Li, L., Keverne, E. B., Aparicio, S. A., Ishino, F., Barton, S. C., and Surani, M. A. (1998). Regulation of maternal behavior and offspring growth by paternally expressed *Peg3*. *Science* **284:** 330–333.
9. Nicholls, R. D. (1998). Imprinting in Prader–Willi and Angelman syndromes. *Trends Genet.* **14:** 194–200.
10. Pengue, G., Calabro, V., Bartoli, P. C., Pagliuca, A., and Lania, L. (1994). Repression of transcriptional activity at a distance by the evolutionarily conserved KRAB domain present in a subfamily of zinc-finger proteins. *Nucleic Acids Res.* **22:** 2908–2914.
11. Peters, J., Wroe, S. F., Wells, C. A., Miller, H. J., Bodle, D., Beechey, C. V., Williamson, C. M., and Kelsey, G. (1999). A cluster of oppositely imprinted transcripts at the Gnas locus in the distal imprinting region of mouse chromosome 2. *Proc. Natl. Acad. Sci. USA* **96:** 3830–3835.
12. Reik, W., and Maher, E. R. (1997). Imprinting in clusters: Lessons from Beckwith–Wiedemann syndrome. *Trends Genet.* **13:** 330–334.
13. Ryan, R. F., Schultz, D. C., Ayyanathan, K., Singh, P. B., Friedman, J. R., Fredericks, W. J., and Rauscher, F. J., III (1999). *KAP-1* corepressor protein interacts and colocalizes with heterochromatic and euchromatic *HP1* proteins: A potential role for Kruppel-associated box-zinc finger proteins in heterochromatin-mediated gene silencing. *Mol. Cell. Biol.* **19:** 4366–4378.
14. Shannon, M., Kim, J., Ashworth, L., Branscomb, E., and Stubbs, L. (1998). Tandem zinc-finger gene families in mammals: Insights and unanswered questions. *DNA Seq.* **8:** 303–315.