

# Retrotransposons Regulate Host Genes in Mouse Oocytes and Preimplantation Embryos

Anne E. Peaston,<sup>1,3,4</sup> Alexei V. Evsikov,<sup>1,3</sup>  
Joel H. Graber,<sup>1</sup> Wilhelmine N. de Vries,<sup>1</sup>  
Andrea E. Holbrook,<sup>1</sup> Davor Solter,<sup>2</sup>  
and Barbara B. Knowles<sup>1,\*</sup>

<sup>1</sup>The Jackson Laboratory  
Bar Harbor, Maine 04609

<sup>2</sup>Max-Planck-Institut für Immunobiologie  
Stübeweg 51  
79108 Freiburg  
Germany

## Summary

**A comprehensive analysis of transposable element (TE) expression in mammalian full-grown oocytes reveals that LTR class III retrotransposons make an unexpectedly high contribution to the maternal mRNA pool, which persists in cleavage stage embryos. The most abundant transcripts in the mouse oocyte are from the mouse transcript (MT) retrotransposon family, and expression of this and other TE families is developmentally regulated. Furthermore, TEs act as alternative promoters and first exons for a subset of host genes, regulating their expression in full-grown oocytes and cleavage stage embryos. To our knowledge, this is the first example of TEs initiating synchronous, developmentally regulated expression of multiple genes in mammals. We propose that differential TE expression triggers sequential reprogramming of the embryonic genome during the oocyte to embryo transition and in preimplantation embryos.**

## Introduction

The growth phase of mammalian oogenesis is characterized by mRNA synthesis and accumulation of stored mRNAs and proteins in the ooplasm. Transcription becomes undetectable in the mouse full-grown oocyte (FGO), restarting in the late zygote, with progressive embryonic genome activation continuing to the morula stage (Latham and Schultz, 2001). During the period of transcriptional silence, maternal mRNAs and proteins, accumulated during oocyte growth, drive meiotic maturation, fertilization, reprogramming of the gametic nuclei, and embryonic genome activation itself. We have previously shown that transposable elements (TEs) are highly expressed during the mouse embryonic genome activation at the 2-cell stage (Evsikov et al., 2004). However, the functional significance of this expression remains unclear (Kigami et al., 2003; Pittoggi et al., 2003).

TEs are a significant component of eukaryotic genomes, occupying more than one third of mouse and

human genomes (International Human Genome Sequencing Consortium, 2001; Mouse Genome Sequencing Consortium, 2002). Almost all mammalian TEs are retrotransposons, which propagate through an RNA intermediate. While the majority are long interspersed nuclear elements (LINEs) or short interspersed nuclear elements (SINEs), approximately one-tenth of them are long terminal repeat (LTR) elements, which resemble retroviruses (Mouse Genome Sequencing Consortium, 2002). Effective propagation of TEs demands their expression in germ cell lineage-competent cells. Indeed, expression of some retrotransposons has been documented in mouse germ cells, gametes, and preimplantation embryos (Bulet et al., 1985; Evsikov et al., 2004; Kigami et al., 2003; Norton and Hogan, 1988; Ostertag et al., 2002; Piko et al., 1984; Poznanski and Calarco, 1991; Taylor and Piko, 1987).

Barbara McClintock documented evidence for a gene regulatory function of TEs in maize, referring to them as controlling units (McClintock, 1950, 1953). Observations of high TE expression in sea urchin and frog eggs and developmental modulation of TE expression suggested the idea that TEs might be used by host organisms as regulatory elements in gene networks (Britten and Davidson, 1969; Davidson and Posakony, 1982). Subsequent genomic analyses and experimental evidence demonstrated that TEs can function as regulatory units for host genes and appear to contribute to many mammalian gene regulatory sequences (Britten, 1997; Jordan et al., 2003; Speek, 2001; van de Lagemaat et al., 2003).

The effects of TEs within their host genome can be viewed as essentially neutral, although their proven capacity for causing deleterious genomic rearrangements and the emergence of mechanisms for their silencing in somatic tissues has led many to view them as harmful parasites (Doolittle and Sapienza, 1980; Hickey, 1982; Orgel and Crick, 1980; Walsh and Bestor, 1999). Extensive genomic alteration by TEs and the functional subversion of many to host physiology prompted the view that TEs exert a powerful influence on genome evolution (Miller et al., 1999; Mouse Genome Sequencing Consortium, 2002; Smit, 1999). TEs influence higher order eukaryotic genomic structure and function by initiating DNA modifications and chromatin remodeling through mechanisms such as RNA interference (RNAi)-mediated TE posttranscriptional processing and DNA targeting (Schramke and Allshire, 2003).

Comparative expression levels of TEs in mammalian oocytes and preimplantation embryos have not been determined, and the range of activated TEs in these unique cells is largely unknown. Analysis of a newly sequenced FGO cDNA library led us to focus on the previously unknown fact that a significant proportion of the maternal mRNA pool is comprised of TEs and chimeric transcripts of TEs with host genes. Here, we report expression of TEs in mouse full-grown oocytes, cleavage stage embryos, and blastocysts. We focus in detail on LTR class III retrotransposons, which include

\*Correspondence: [bbk@jax.org](mailto:bbk@jax.org)

<sup>3</sup>These authors contributed equally to this work.

<sup>4</sup>Present address: School of Molecular and Microbial Biosciences, The University of Sydney, NSW 2006, Australia.

Table 1. Expression of Transposable Elements in Full-Grown Oocyte, 2-Cell Stage Embryo, and Blastocyst cDNA Libraries

Repetitive Element	Number of ESTs in the Library (% of All ESTs)					
	Full-Grown Oocyte		2-Cell		Blastocyst	
MT	2403	(12.73)	280	(2)	1	(0.01)
ORR	0		108	(0.77)	6	(0.04)
MuERV-L	0		443	(3.16)	0	
IAP (ERV-K)	6	(0.14)	79	(0.56)	35	(0.23)
LINE	8	(0.04)	53	(0.38)	22	(0.15)
Other <sup>a</sup>	13	(0.07)	96	(0.68)	15	(0.1)

<sup>a</sup> SINE elements were excluded from this analysis because size selection of transcripts for library construction prevented their representative inclusion. They are estimated to comprise 3%–5% of polyadenylated mRNA in oocytes and a lower proportion in blastocysts (Bachvarova, 1988).

members of the mammalian apparent LTR retrotransposon (MaLR) and endogenous retrovirus L (ERV-L) families, and we investigate the ability of TEs to act as stage-specific alternative promoters for a number of host genes.

## Results

### LTR Class III Retrotransposons Are Preferentially Expressed in Full-Grown Oocytes and Cleavage Stage Embryos

To determine the overall pattern of TE expression in full-grown oocytes and preimplantation embryos, we analyzed the number of repetitive element ESTs in large, representative, nonnormalized cDNA libraries from full-grown oocytes and 2-cell stage embryos and blastocysts (Table 1; Rothstein et al., 1992). MT (mouse transcript), a member of the MaLR family of nonautonomous retrotransposons, accounted for over 12% of the total ESTs in the full-grown oocyte cDNA library. Other TEs contributed collectively less than 0.3% of ESTs. In contrast to the oocyte, the bulk of interspersed repeat ESTs in the 2-cell stage embryo cDNA library were murine ERV-L (MuERV-L) (Benit et al., 1997; Evsikov et al., 2004). MT and MuERV-L were in very low abundance in the blastocyst cDNA library (Table 1).

Expression of representatives from all three classes of LTR retrotransposons found in FGO and/or 2-cell cDNA libraries was analyzed in more detail, using reverse transcription polymerase chain reaction (RT-PCR). RLTR1B (LTR class I), IAPEz (LTR class II), and ORR1A1, MT, and MuERV-L (all LTR class III) showed different patterns of expression (Figure 1). Interestingly, antisense transcripts of MuERV-L were also coexpressed in a stage-specific pattern with sense counterparts, potentially setting the stage for formation of double-stranded RNA and subsequent RNAi-dependent MuERV-L silencing.

### Transposable Elements Provide an Alternative 5' Exon to Many Transcripts in Full-Grown Oocytes and Cleavage Stage Embryos

In addition to the TEs themselves, the FGO and 2-cell cDNA libraries contain many chimeric gene transcripts characterized by alternative TE-derived 5' sequence. Except for this 5' sequence, such transcripts are identical to mRNA sequences of known host genes (i.e., mouse genes curated by Mouse Genome Informatics [MGI] and/or NCBI Gene databases). These chimeric

transcripts constitute 3% and 1.4% of the 3000 most abundantly represented, identified gene transcripts in the FGO and 2-cell stage embryo cDNA libraries, respectively (Supplemental Tables S2–S4 at <http://www.developmentalcell.com/cgi/content/full/7/4/597/DC1/>). The listed number of genes that have chimeric mRNAs is likely conservative, given that many transcripts in oligo-dT primed cDNA libraries are 5'-truncated.

To address whether chimeric transcripts belong to genes of any particular functional category, gene ontology (GO) annotations of the genes were analyzed using the MGI GO Slim Chart Tool ([http://www.spatial.maine.edu/~mdolan/MGI\\_GO\\_Slim\\_Chart.html](http://www.spatial.maine.edu/~mdolan/MGI_GO_Slim_Chart.html)). Although approximately 50% of genes have yet to be annotated, no

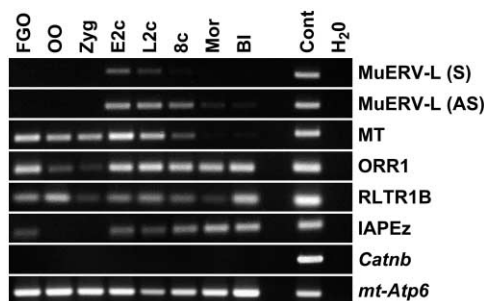


Figure 1. Expression of LTR Retrotransposons in Full-Grown Oocytes and Preimplantation Embryos

Retrotransposons are identified to the right of each panel by symbol: MuERV-L (S) and MuERV-L (AS), sense and antisense transcripts of murine endogenous retrovirus L, respectively; MT, mouse transcript nonautonomous retrotransposon; ORR1, Origin-Region Repeat 1 nonautonomous retrotransposon (Smit, 1993); RLTR1B, subfamily of ERV1 (Jurka, 2000); IAPEz, Intracisternal A-type Particle provirus, ERV-K family (Jurka, 2000). Bottom two panels, controls: *Catnb*, intronic region of  $\beta$ -catenin gene (control for genomic DNA contamination); *mt-Atp6*, mitochondrial ATP synthase 6 (internal control of cDNA quality).

Abbreviations: OO, ovulated oocyte; Zyg, zygote (1-cell embryo); E2c, early 2-cell stage embryo; L2c, late 2-cell stage embryo; 8c, early 8-cell stage embryo; Mor, morula; Bl, blastocyst; Cont, positive control for PCR reactions (C57BL/6J genomic DNA for retrotransposons and *Catnb*; plasmid with mitochondrial ATP synthase 6 insert for *mt-Atp6*); H<sub>2</sub>O, water control for PCR reactions.

PCR results were the same using two different sets of primers for ORR1 and IAPEz (Supplemental Table S5); for MT and MuERV-L, different primer sets were used to those described previously (Evsikov et al., 2004). A representative PCR result is shown.

Table 2. Classification of Chimeric Transcripts in Full-Grown Oocyte and 2-Cell Stage Embryo cDNA Libraries According to the Contributing Retrotransposon

Repeat Class	LTR			Number of Genes		
	Class	Family	Element or Subfamily	FGO	2-Cell	
DNA				1	0	
LINE				3	1	
SINE				11	8	
LTR	I	ERV1	LTRIS_MM	1	0	
			MMURS	1	0	
			RMER	RMER6, 15, 19A, 19B	5	1
	II	ERV-K	RMER4b, RLTR10, IAPLTR2_MM, BGLII, RLTR9A, ETnERV2	8	1	
	III	ERV-L	MuERV-L	3	10	
			MT2B	3	9	
			MT2C	0	4	
			MaLR	MT	49	7
			ORR1	5	0	
			MLT	1	0	
Complex				5	0	
Total				96	41	

TE-derived transcripts from the most abundantly represented 3000 named genes in the FGO and 2-cell embryo cDNA libraries, classified according to the repetitive element class from which the 5' sequence was derived. Complex 5' sequences, derived from more than one TE, all include a MaLR retrotransposon as one component.

bias toward any particular category was revealed (data not shown).

To determine which TEs are most frequently involved in forming chimeric transcripts, they were classified according to the TE contributing an alternative 5' sequence (Table 2). In oocytes, MaLR retrotransposons and notably MT were the primary contributors to alternative 5' sequences. MTA, phylogenetically the youngest and most abundant MT subfamily (Smit, 1993), contributed 5' sequence to 27% of chimeric transcripts (Supplemental Tables S2 and S3). In the 2-cell stage embryo, MuERV-L and other rodent ERV-L elements (MT2B, MT2C) together contributed 5' sequence to 56% of chimeric transcripts. Therefore, TE composition of the chimeric transcripts correlates with the overall abundance of a given transposable element in the oocyte and 2-cell stage embryo libraries. Interestingly, a greater variety of TEs contribute to chimeric transcripts in the full-grown oocyte than in the 2-cell stage embryo.

#### Origin of Chimeric Transcripts

To gain insight into the mechanism(s) of chimeric transcript generation, sequences were aligned to the annotated Ensembl mouse genome assembly using SSAHA (sequence search and alignment by Hashing algorithm) (Ning et al., 2001). In all cases, the cognate TEs were located within the cellular gene locus or upstream of it. Chimeric transcripts were missing all the exons located upstream of the TEs and usually lacked one or more conventional 5' exons when the TE was located upstream of the gene locus (Table 3).

BLASTN searches of the expressed sequence tag database (dbEST) (Boguski et al., 1993) with these chimeric transcripts were used to determine (1) whether there was evidence that such transcripts were splice isoforms, as opposed to transcripts arising from an alternative promoter, and (2) whether such transcripts could be

found in other cDNA libraries. All matching ESTs incorporated the TE as the 5' terminal sequence, not as an internal sequence. Moreover, all such ESTs were found in various ovary, oocyte, or preimplantation embryo cDNA libraries, but never in cDNA libraries from other developmental stages or tissues.

#### Conserved LTR Splice Donor Site with Gene AT-Rich Splice Acceptor Site Characterize MTA LTR-Derived Chimeric Transcripts

A neural network splice site predictor (Reese et al., 1997) was used to determine how MaLR LTRs might splice to cellular genes. The MaLR LTR of chimeric transcripts was always found in its sense orientation. Analysis of consensus MaLR LTR sense sequences predicted a conserved potential splice donor site immediately upstream of the LTR polyadenylation site of many MaLRs (Figure 2A). Splice site analysis of ten specific MTAs generating a chimeric transcript showed that in most cases, the predicted site matched the actual one (Figure 2B). Despite the splicing activity of some, most MT ESTs in the FGO cDNA library do not form chimeric transcripts but are clearly the transcripts of cognate unspliced full-length MT retrotransposons.

Bias against orientation of the TE in the sense orientation, relative to the transcription unit in which it is found, is slight for MTs, suggesting this is not an important factor for MTA splicing activity (Figure 2C). In contrast, genomic analysis showed that a markedly AT-rich sequence of at least 100 nt characterizes the region immediately 5' of the acceptor site for actively splicing MTAs (Figure 2D).

Taken together, the data suggest that a conserved MTA LTR-splice site functionally competes with the predicted LTR polyadenylation and cleavage site and may be preferentially used when the MTA is inserted and expressed in a suitable genomic location. Bias against

Table 3. Intron/Exon Structure of Selected Chimeric Transcripts

Transposable element	Symbol	Schematic	
MTA	<i>Stk3</i>		
	<i>5730494M16Rik</i>		
	<i>Speer5-pps1</i>		
	<i>Zfp277</i>		
	<i>Spin</i>		
	<i>D7Ertd445e</i>		
	<i>Dncic2</i>		
	<i>D6Ertd365e</i>		
	<i>C230040D10Rik</i>		
	<i>Abcb1b</i>		
	<i>Fert2</i>		
	<i>Ski</i>		
	MTC	<i>Vdac2</i>	
		<i>Nfil3</i>	
<i>Dnajc11</i>			
<i>Trp53bp1</i>			
MTD		<i>D6Ertd527e</i>	
	<i>AU017455</i>		
	<i>Rnf24</i>		
RLTR10/MTE	<i>Pard3</i>		
MT-int (MTA)	<i>Calr3</i>		
MT-int (MTC)	<i>E330021D16Rik</i>		
MT2B	<i>2610005H11Rik</i>		

Examples of chimeric transcript structure determined by alignment of the transcript to the Ensembl annotated mouse genome assembly, release 13.30.1. The retrotransposon alternative first exon (red box) is shown in relation to the contiguous gene; white boxes – conventional transcript exons omitted in chimeric transcript; black boxes, conventional transcript exons included in chimeric transcript.

fixation of intronically located MaLRs in the same orientation as their human or mouse host gene is not uniform between families. Our data suggest that factors other than LTR polyadenylation and cleavage sites affect orientation of LTR retrotransposons within genes (Medstrand et al., 2002; Smit, 1993).

### MT LTRs and Developmental Regulation of Chimeric Transcript Expression

To investigate expression of MT-derived chimeric transcripts relative to their conventional nonchimeric counterparts, RT-PCR analysis of specific transcript expression in FGOs and preimplantation embryos was performed.

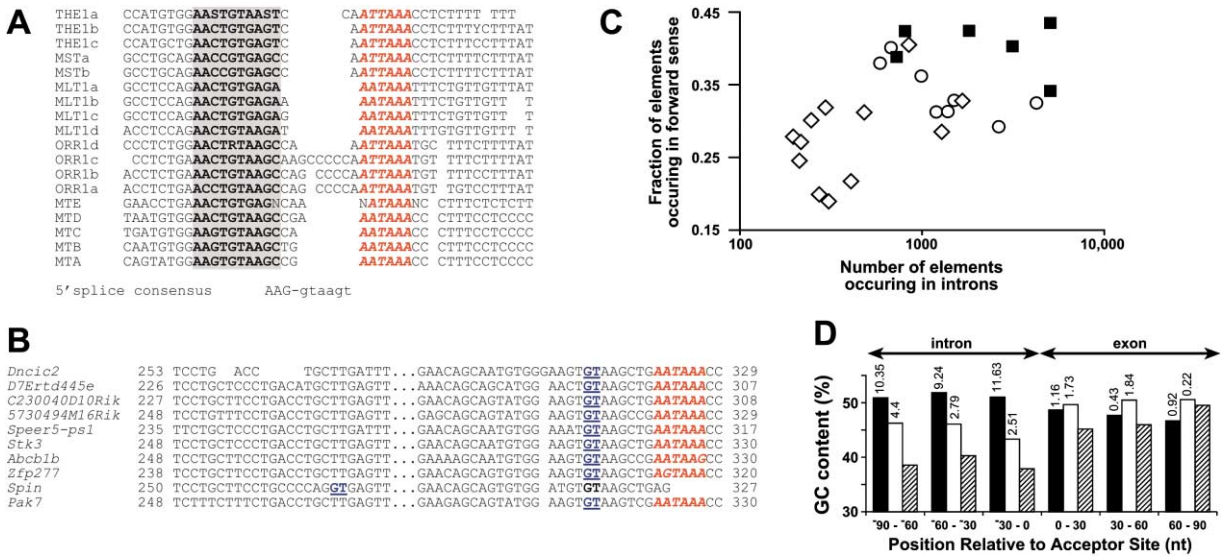


Figure 2. Structural Analyses of MTA LTRs and of MaLR Genomic Positioning

(A) Aligned consensus MaLR LTRs from RepBase (Jurka, 2000) compared with a consensus mammalian splice site (Padgett et al., 1986). Bold shaded sequence: putative conservation of 5' splice site (donor). Red bold italics: putative polyadenylation site. S: G or C. Lower case letters: intron sequence.

(B) Alignment of 10 MTAs that contribute 5' sequence to chimeric transcripts. Bold: predicted splice sites; blue underline: actual splice sites; red bold italics: putative MTA LTR polyadenylation site. The upstream splice site in *Spin* is a result of T→G transversion, which creates a strong predicted donor site. Splicing is processed 5'→3' (Staley and Guthrie, 1998), so this site, occurring first, is preferentially used.

(C) Orientation bias of MaLR elements within introns. The orientation of all MTs occurring within introns in the mouse genome was determined and compared with that of other intronically located MaLRs (ORR1 and MLT). Y axis: fraction of indicated intronic MaLR elements that have the same orientation as the surrounding transcriptional unit. X axis: number of the indicated MaLR elements occurring in introns. Black squares, MT subfamilies; circles, ORR1 subfamilies; diamonds, MLT subfamilies. No obvious correlation between the degree of bias against same orientation and the strength of the prediction for the splice site of the consensus sequence was found for any element (data not shown).

(D) MTA LTRs splice into acceptor sites with a high upstream AT content. The nt composition of 180 nt window flanking the intron/exon boundary was compared between (1) MTA elements spliced to a downstream exon, (hatched columns; n = 24); (2) the acceptor site 3' of the MTA LTR in any Ensembl gene containing an MTA LTR in the same orientation as the transcript (black columns; n = 8287); (3) a pseudo-random sampling of acceptor sites from any Ensembl transcript (gray columns; n = 12,043). X axis: distance from the splice junction in 30 nt windows. Y axis: mean GC content in sample population. The numbers above the columns are the Z values associated with the test for equality of GC-content between the active MTA LTR (hatched column) and the sets in the other columns. The greater the Z value, the greater the difference between sets.

Chimeric and nonchimeric forms of *Spin* (spindlin), *Dnajc11* (DnaJ (Hsp40) homolog, subfamily C, member 11), and *Vdac2* (voltage-dependent anion channel 2) were detected in oocytes and early preimplantation embryos but exhibited different patterns of expression (Figure 3A). In contrast, only the chimeric form of *Nfil3* (nuclear factor, interleukin 3, regulated) was expressed in oocytes and early embryos. Interestingly, both *Nfil3* chimeric and conventional transcripts encode identical predicted proteins. Disappearance of chimeric transcripts by the 8-cell embryo stage suggests that they are maternal mRNAs and that MTs may act as oocyte-specific promoters. Moreover, regulation of MT-derived chimeric transcripts is different from conventional transcriptional control of respective host genes.

To determine whether MTA-derived chimeric transcripts might be regulated by another promoter, we used the Gibbs Sampler (Lawrence et al., 1993) to search for conserved DNA patterns in the 1 kb genomic sequence immediately upstream of MTA LTRs. No patterns considered statistically significant ( $p > 0.1$ , Wilcoxon Test) were found, adding further evidence to the notion that the *cis* regulatory element, which controls chimeric transcript expression, is most likely the MT LTR.

### MT-Derived SPIN Protein in 2-Cell Stage Mouse Embryos

One-third of the chimeric transcripts from the FGO have predicted open reading frames identical to the nonchimeric transcript isoforms, whereas others presumably encode alternative polypeptides (Supplemental Tables S2 and S3). To determine whether chimeric transcript-encoded protein isoforms are translated, we prepared antibody to the N-terminal oligopeptide of the chimeric form of *Spin* (Figure 3C). This antibody immunoprecipitated a single band, of the correct size for chimeric SPIN protein, from <sup>35</sup>S-methionine-labeled 2-cell stage embryos (Figure 3B). SPIN is an abundant maternal protein that is phosphorylated in the MAP kinase pathway (Oh et al., 1997, 1999). The predicted N termini of the chimeric and conventional SPIN protein isoforms have different potential phosphorylation sites (Figure 3C), suggesting that these two isoforms may function differently.

### Phylogenetic Conservation

To approach the question of functional significance from a different angle, class III LTR retrotransposons that

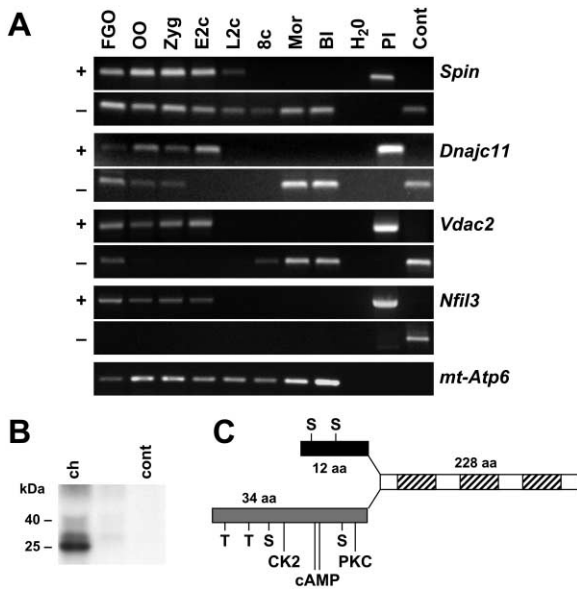


Figure 3. Expression of Chimeric Transcripts Driven by MT Retrotransposons

(A) RT-PCR of chimeric (+) or conventional (-) transcripts in FGOs and preimplantation embryos. Amplicon band size and sequencing confirmed the identity and structure of chimeric transcripts. Transcripts are identified by gene symbol to the right of the chimeric panel of each set. Lanes identified as in Figure 1. Control: R1 ES cell cDNA (except *Nfil3*, mouse brainstem cDNA); PI, control plasmids with chimeric transcript inserts; *mt-Atp6* typical example is in the lowest panel, a plasmid control is not included in this example. (B) Autoradiograph showing translation of chimeric *Spin* transcript. Ch: immunoprecipitation using affinity-purified rabbit polyclonal antibody against N-terminal polypeptide of chimeric SPIN protein; Cont: rabbit preimmune serum. Expected protein size is 27 kDa. (C) *Spin* protein product schematic showing alternative N termini with different predicted phosphorylation sites. Black bar, MT-encoded amino acids; gray bar, amino acids encoded by conventional nonchimeric transcript; white bar, amino acids common to both protein isoforms; striped boxes in white bar, SSTY motifs characteristic of *Spin-Ssty* family.

contribute to specific chimeric transcripts were assessed for their evolutionary conservation. Alignment of chimeric transcripts to 129X1/SvJ mouse genome sequence (Celera Discovery System and Celera's associated databases) revealed that individual TEs, including mouse-specific repeats such as MTA and MuERV-L, were conserved in the respective loci of two distantly related strains of inbred mice (C57BL/6J and 129X1/SvJ). Positional conservation of older rodent TEs in a number of genes was determined by analyzing the syntenic rat gene locus (Ensembl rat genome assembly 2). For some of these genes, chimeric and conventional transcripts encode identical proteins (e.g., *Nfil3* [MTC], *2610005H11Rik* [MT2B], *Rpl41* [ribosomal protein L41, MT2B], and *Rpl17* [ribosomal protein L17, RMER6A]). However, some of the conserved elements give rise to chimeric transcripts that may encode altered proteins, e.g., *Vdac2* (MTC), *Itpr5* (inositol 1,4,5-triphosphate receptor 2, MTC), *Eef2k* (eukaryotic elongation factor-2 kinase, RMER15).

Interestingly, different MT2B elements display dissimilarities in transcriptional control of different genes.

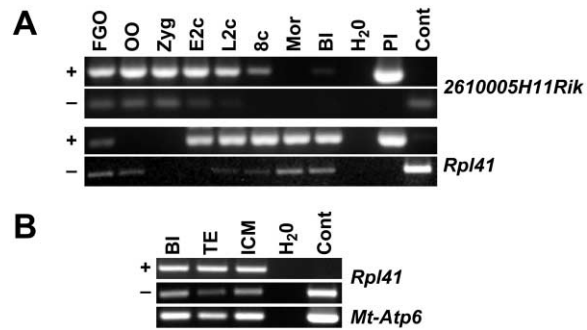


Figure 4. Expression of Chimeric Transcripts Driven by MT2B Retrotransposons

(A) RT-PCR of chimeric (+) or conventional (-) transcripts in FGOs and preimplantation embryos. Labeling as in Figure 3A. Control: mouse brainstem cDNA.

(B) RT-PCR of chimeric (+) or conventional (-) transcripts of *Rpl41* gene in blastocysts (Blast) and isolated trophectoderm (TE) and inner cell mass (ICM). *mt-Atp6*, mitochondrial ATP synthase 6 (internal control of cDNA quality); H<sub>2</sub>O, water control for PCR reactions; Control: mouse liver cDNA (*Rpl41*), plasmid (*mt-Atp6*).

MT2B-driven expression of *Rpl41* gene is reactivated at the time of embryonic genome activation and persists at least into the blastocyst. Expression of the *2610005H11Rik* chimeric isoform is undetectable after the 8-cell stage, suggesting that the MT2B promoter of this gene is active only in oocytes (Figure 4A). We examined whether the first differentiation event in mouse embryos, formation of trophectoderm at the blastocyst stage, may result in the change of transcriptional regulation of chimeric transcripts. We did not observe a difference in expression of chimeric and nonchimeric forms of *Rpl41*, both of which are transcribed at this stage, in trophectoderm and inner cell mass (Figure 4B). Therefore, the general mechanism of MT2B-driven transcriptional regulation may still be retained in the cells after commitment to the trophectoderm lineage.

## Discussion

Analysis of the full-grown oocyte transcriptome reveals the high contribution of transposable elements to the maternal mRNA pool. Different LTR retrotransposons have specific, developmentally regulated expression patterns, suggesting that normal repressive chromatin structure for these loci is established sequentially during the oocyte-to-embryo transition and preimplantation stages. Moreover, LTR retrotransposons in particular, and occasionally other TEs, act as oocyte- and preimplantation embryo-specific alternative promoters for a wide variety of host genes. In these alternative transcripts, TEs contribute an alternative 5' exon. This introduces variation in gene expression and potentially alters gene function either at the RNA or protein level. The list of chimeric transcripts presented here greatly expands the number of known genes with alternative TE-derived promoters (van de Lagemaat et al., 2003). However, these transcripts are found only in a very specific time period.

### TEs Modify Gene Products and Potentially Initiate New Gene Regulatory Systems in Oocytes and Early Embryos

The synchronous expression of multiple genes driven by TEs is an example of how random insertions of regulatory elements, such as the MT LTR, can result in coregulated gene expression. Functional proteins may be produced by chimeric transcripts with the same coding sequence as their conventional counterparts (e.g., *Nfil3*, *Rpl41*, *2610005H11Rik*). Chimeric transcripts may also produce nonconventional proteins, as shown by the variant SPIN protein. Direct determination of the function of these variant chimeric transcripts requires investigation of each transcript and its cognate protein as the oocyte progresses through maturation, fertilization, and activation of the embryonic genome.

TE-driven transcription of multiple host genes described here provides, in principle, rich grounds for selection of new modes of gene regulation by introducing substantial variation in gene expression and possibly function. In mammals, such selection of new variants at the oocyte and early embryo stages may be feasible, since it would not affect the fitness of females (although it may affect reproductive success). Some of the new TE-induced modes of gene regulation may, with time, get adopted and even gain different regulatory specificity, such as activation of TE-driven transcription at other times of development. For example, the MT2B promoter of *Rpl41* regulates expression of this gene in preimplantation embryos, as well as oocytes, whereas the MT2B promoter of the *2610005H11Rik* gene is active only in oocytes.

### Oocytes, Embryos, and the Biology of MaLR Retrotransposons

The current study indicates that oocytes and cleavage stage embryos provide an environment particularly suitable for transcription of the MaLR and ERV-L families of LTR class III retrotransposons and relatively less suitable for transcription of class I and II LTR retrotransposons and LINEs. MaLRs encode no known proteins, and their means of propagation in the genome is unknown. We have shown that reverse transcriptase activity is extremely high in the early cleavage stage embryos, when MuERV-L expression is maximal (Evsikov et al., 2004). The abundance of MT and ORR mRNAs during the same stages (Table 1, Figure 1) prompts the suggestion that these retrotransposons may use MuERV-L reverse transcriptase to propagate.

MaLRs are the most numerous LTR family in the mouse genome, with approximately 388,000 copies (Mouse Genome Sequencing Consortium, 2002). They have coevolved with mammalian genomes for over 150 million years and, in the case of MT and ORR subfamilies, with rodent genomes for the last 75 million years (Smit, 1996). MTA may still be active in the mouse genome (Smit, 1993); indeed, one instance of MTA insertional pathology that results in structural alteration of the affected transcript has been reported (Loftus et al., 1997). Results of this study, when combined with those from our previous work (Evsikov et al., 2004), suggest that new class III LTR retrotransposon insertions into the host genome may occur in the cleavage stage embryo.

### Mechanisms of Differential TE Expression in the Oocyte and Preimplantation Embryo

Transcription of mobile elements within any given cell type depends on the availability of TE-competent transcription machinery and/or the epigenetic status of the genomic locus of each element. Thus, differential expression of TEs may in part be regulated by the changing complement of transcription factors in growing oocytes compared to cleavage-stage embryos (Wang et al., 2004). Transcriptional specificity of LTR retrotransposons is conferred by multiple transcription factor binding sites in LTR sequences and can vary considerably within and between different classes (Keshet et al., 1991). Factors relevant to transcription of MaLRs and MuERV-L are currently completely unknown.

Epigenetic changes to the TE loci may also cause repression/derepression of each different class of mobile elements. For example, cytosine methylation, which is associated with silencing of TEs and imprinted genes (Yoder et al., 1997), occurs unequally between the maternal and paternal genomes in the zygote (Arney et al., 2002; Mayer et al., 2000; Oswald et al., 2000). Gradual genome-wide demethylation commences after the first cleavage, and remethylation begins in the blastocyst (Monk et al., 1987; Santos et al., 2002). Therefore, transcription profiles of class III LTR retrotransposons do not perfectly coincide with global cytosine methylation changes. This may be reflective of their unequal methylation in the maternal and paternal genomes and/or their remethylation in the early, rather than late, preimplantation embryo.

The RNA interference machinery has been documented to silence retrotransposons in both plants and animals and is required for DNA methylation in most eukaryotes studied (Freitag et al., 2004; Hannon, 2002). In mouse FGOs and early embryos, components of the RNAi machinery are present and active (Evsikov et al., 2004; Svoboda et al., 2000). In addition, both sense and antisense transcripts of MuERV-L and other TEs are coexpressed, which may enable formation of double-stranded RNA and trigger RNAi. Other indirect evidence also supports the notion that MuERV-L and IAP may be regulated by this mechanism (Svoboda et al., 2004). An RNAi-dependent mechanism is involved in heterochromatin formation in fission yeast and *Drosophila*. In mammalian cells, pericentromeric heterochromatin and retrotransposons associate with similar chromatin remodeling factors (Hakimi et al., 2002; Kondo and Issa, 2003; Schramke and Allshire, 2003; Volpe et al., 2002). Taken together, these facts raise the possibility that serial activation and silencing of retrotransposons observed in oocytes and preimplantation embryos may reflect stage-specific, potentially RNAi-mediated, targeting of chromatin remodeling complexes to widely distributed regions of the embryonic genome.

### Concluding Remarks

A substantial cohort of host genes described here, whose developmentally regulated expression is controlled by TE-derived 5' exons, demonstrates that retrotransposons may directly affect developmental processes in oocytes and cleavage stage embryos. Such an abundance of chimeric transcripts is not detectable

by microarray analysis of mRNA expression in oocytes and early embryos. Most of the array probes used in such analyses are 3' biased (Carter et al., 2003; Hamatani et al., 2004; Wang et al., 2004; Affymetrix Technical Note Part No. 701405), making it impossible to discern the transcripts with alternative 5' ends. Even if 5' array probes are used, the repetitive nature of these alternative 5' exons would likely cause nonspecific hybridization, precluding unambiguous results for such transcripts. Microarray analysis of chimeric transcripts is technically possible, e.g., using techniques for analysis of alternative transcripts (Hu et al., 2001), but requires initial knowledge of TE-derived transcripts in the sample. Therefore, study of the transcriptome through direct sequence analysis is a very important tool for discovery in the oocyte to embryo transition.

Sequential activation and silencing of MaLR and other retrotransposons in oocytes and preimplantation embryos, together with the known links between retrotransposon silencing and chromatin remodeling, lead us to propose that genome remodeling during this period could be initiated and ordered by retrotransposon expression. Concurrent analysis of both transcriptional activation and epigenetic modifications of specific genomic loci will be required to explore this idea.

#### Experimental Procedures

##### Analysis of cDNA Libraries and Chimeric Transcripts

mRNA from full-grown germinal vesicle stage oocytes, 2-cell stage embryos, and blastocysts was used to construct oligo-dT-primed cDNA libraries as described (Rothstein et al., 1992). Analysis of ESTs from the 2-cell stage embryo cDNA library (14,813 ESTs; dbEST library ID.862) was reported (Evsikov et al., 2004). ESTs from the FGO library (19,000 ESTs; I.M.A.G.E. library ID, 1182) and two blastocyst cDNA libraries (15,454 ESTs; dbEST library ID.850 and ID.875) were analyzed similarly. In brief, overlapping ESTs were assembled into consensus sequences. Each consensus sequence or single EST represented an individual transcriptional unit; the number of ESTs per unit indicated its relative abundance in the library. BLASTN (Altschul et al., 1990) searches of GenBank and SSAHA (Ning et al., 2001) searches of the Ensembl mouse annotated genome assembly (<http://www.ensembl.org>), release 13.3.01, were performed, in combination with analysis by RepeatMasker version 07/07/01 (<http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker>) using each consensus sequence or single EST to identify genes and classify repeats. For transcripts with repeat-derived alternative 5' sequences, final repeat identity was assigned by alignment of the transcript to the annotated Ensembl public mouse genome assembly, Build 30, annotated using RepeatMasker based on the March 2002 Repbase Update (Jurka, 2000).

To determine potential open reading frames (ORFs) encoded by chimeric transcripts, we used ORF finder (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>). The longest ORF starting with ATG was then aligned to GenBank protein sequences using BLASTP. Chimeric transcript-derived ORFs were defined as no different from the conventional protein if the N-terminal and all downstream sequences were identical to the conventional protein.

##### Oocyte and Embryo Preparation

All mice used in this study were B6D2F1/J. Oocytes and embryos were isolated as described (Oh et al., 1999, 2000). For RT-PCR assays, batches of oocytes or embryos (55 per batch) were harvested, processed as described, snap frozen in liquid nitrogen, and stored at  $-80^{\circ}\text{C}$  until use (Evsikov et al., 2004). Inner cell masses of blastocysts were isolated by immunosurgery as described (Solter and Knowles, 1975). Trophectoderms of expanded blastocysts were biopsied using a DIC-equipped Nikon inverted microscope and Narishige micromanipulators.

##### cDNA Preparation and Analysis

Bacterial clones bearing plasmids containing inserts of interest were obtained either from the FGO library or American Type Culture Collection (I.M.A.G.E. clones). Plasmid inserts were sequenced and verified by comparison with the cluster sequence from either FGO cDNA library or GenBank. Optimal PCR conditions for each PCR primer pair were determined empirically.

RNA was isolated as described (Oh et al., 2000) or using PicoPure RNA Isolation Kit (Arcturus, Mountain View, CA). DNase-treated RNA was used for cDNA first-strand synthesis (Superscript Preamplification System, Invitrogen). Absence of genomic DNA contamination was confirmed by PCR of a  $\beta$ -catenin (*Catnb*) intronic region. PCR of mitochondrial ATP synthase 6 (*mt-Atp6*) was used as a constitutive control. PCRs analyzing chimeric transcripts were conducted twice on at least two independent sets of random hexamer- or oligo-dT-primed cDNA; PCRs analyzing TE expression used at least two independent sets of oligo-dT-primed cDNA. The cDNA template in each PCR reaction was two oocytes or embryo equivalents. Under our PCR conditions, we performed between 30 and 40 cycles of amplification to optimize detection of rare transcripts. Primers for chimeric and nonchimeric transcript PCRs spanned more than one exon; primers for LTR TEs were designed to amplify the internal region of the element (Supplemental Table S5).

##### Antibody Preparation and Immunoprecipitation

A peptide corresponding to the N terminus of protein encoded by *Spin* chimeric transcript (MASASSPASSPRK, 12 aa) was synthesized and used in the preparation of polyclonal antiserum in rabbits (Biosource International, Hopkinton, MA). Affinity-purified serum was used to immunoprecipitate protein from 100  $^{35}\text{S}$ -methionine-labeled 2-cell stage mouse embryos, as described previously (Oh et al., 1997, 2000). Eluted protein was separated by SDS-PAGE and autoradiographs prepared. Bands were identified by comparison with protein size markers separated on the same gel.

##### Computational Analyses

We used neural network software, NNSPLICE (Reese et al., 1997) 0.9 version, to predict 5'-splice sites in MaLR consensus sequences and in the specific MTA elements giving rise to chimeric transcripts. Consensus MaLR sequences were from Repbase 8.4 (Jurka, 2000). Actual 5' splice sites of MTs were determined by chimeric transcript alignment with genomic sequence.

To analyze the base composition of 3'-splice (acceptor) sites, the 180 nt flanking the acceptor sites were extracted for: (1) MTA-derived chimeric transcripts from the FGO library; (2) any Ensembl transcript whose genomic precursor contained an intronic MTA LTR in the same transcriptional orientation; the next acceptor site downstream of the MTA was chosen; (3) a pseudo-random sampling of acceptor sites from any Ensembl transcript. The GC-percentage was measured in nonoverlapping 30 nt windows for each sequence set. Equality of the GC-content in different sequence sets in each window was tested using a two-sided, large-sample, standard-normal approximation of the binomial distribution (Miller and Freund, 1965).

To search for potential genomic promoter patterns, 1000 nts of genomic DNA sequence 5' upstream of the MTA for 19 MTA-derived chimeric transcripts were analyzed. DNA patterns common to these sequences were searched using the Gibbs Sampler (Lawrence et al., 1993) with default search parameters and patterns lengths 10, 15, and 20 nt. Statistical significance of patterns was determined with the Wilcoxon test implemented in the Gibbs Sampler.

##### Acknowledgments

We thank Marge May of the The Jackson Laboratory Computational Biological Resources group for assistance with sequence analysis programs and Emily Radford for technical assistance. Dr. Arian Smit of The Institute of Systems Biology kindly participated with us in many helpful discussions. The authors thank Drs. Wayne Frankel, Tatyana Golovkina, Keith Hutchinson, Carrie Marin, and the anonymous reviewers for critical reading and helpful suggestions on the manuscript. This work was supported by NHMRC of Australia CJ Martin Fellowship 007150 (A.E.P.), The Lalor Foundation (A.V.E.),



the Max Planck Society (D.S.), US PHS National Institutes of Health grants RR16463, RR018789, HD37102, and CA34196, and the National Science Foundation grant EPS-0132384.

Received: January 23, 2004

Revised: August 11, 2004

Accepted: August 12, 2004

Published: October 11, 2004

## References

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.
- Arney, K.L., Bao, S., Bannister, A.J., Kouzarides, T., and Surani, M.A. (2002). Histone methylation defines epigenetic asymmetry in the mouse zygote. *Int. J. Dev. Biol.* **46**, 317–320.
- Bachvarova, R. (1988). Small B2 RNAs in mouse oocytes, embryos, and somatic tissues. *Dev. Biol.* **130**, 513–523.
- Benit, L., De Parseval, N., Casella, J.F., Callebaut, I., Cordonnier, A., and Heidmann, T. (1997). Cloning of a new murine endogenous retrovirus, MuERV-L, with strong similarity to the human HERV-L element and with a gag coding sequence closely related to the Fv1 restriction gene. *J. Virol.* **71**, 5652–5657.
- Boguski, M.S., Lowe, T.M., and Tolstoshev, C.M. (1993). dbEST—database for expressed sequence tags. *Nat. Genet.* **4**, 332–333.
- Britten, R.J. (1997). Mobile elements inserted in the distant past have taken on important functions. *Gene* **205**, 177–182.
- Britten, R.J., and Davidson, E.H. (1969). Gene regulation for higher cells: a theory. *Science* **165**, 349–357.
- Brulet, P., Condamine, H., and Jacob, F. (1985). Spatial distribution of transcripts of the long repeated ETn sequence during early mouse embryogenesis. *Proc. Natl. Acad. Sci. USA* **82**, 2054–2058.
- Carter, M.G., Hamatani, T., Sharov, A.A., Carmack, C.E., Qian, Y., Aiba, K., Ko, N.T., Dudekula, D.B., Brzoska, P.M., Hwang, S.S., and Ko, M.S. (2003). In situ-synthesized novel microarray optimized for mouse stem cell and early developmental expression profiling. *Genome Res.* **13**, 1011–1021.
- Davidson, E.H., and Posakony, J.W. (1982). Repetitive sequence transcripts in development. *Nature* **297**, 633–635.
- Doolittle, W.F., and Sapienza, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284**, 601–603.
- Evsikov, A.V., de Vries, W.N., Peaston, A.E., Radford, E.E., Fancher, K.S., Chen, F.H., Blake, J.A., Bult, C.J., Latham, K.E., Solter, D., and Knowles, B.B. (2004). Systems biology of the 2-cell mouse embryo. *Cytogenet. Genome Res.* **105**, 240–250.
- Freitag, M., Lee, D.W., Kothe, G.O., Pratt, R.J., Aramayo, R., and Selker, E.U. (2004). DNA methylation is independent of RNA interference in *Neurospora*. *Science* **304**, 1939.
- Hakimi, M.A., Bochar, D.A., Schmiesing, J.A., Dong, Y., Barak, O.G., Speicher, D.W., Yokomori, K., and Shiekhattar, R. (2002). A chromatin remodelling complex that loads cohesin onto human chromosomes. *Nature* **418**, 994–998.
- Hamatani, T., Carter, M.G., Sharov, A.A., and Ko, M.S. (2004). Dynamics of global gene expression changes during mouse preimplantation development. *Dev. Cell* **6**, 117–131.
- Hannon, G.J. (2002). RNA interference. *Nature* **418**, 244–251.
- Hickey, D.A. (1982). Selfish DNA: a sexually-transmitted nuclear parasite. *Genetics* **101**, 519–531.
- Hu, G.K., Madore, S.J., Moldover, B., Jatke, T., Balaban, D., Thomas, J., and Wang, Y. (2001). Predicting splice variant from DNA chip expression data. *Genome Res.* **11**, 1237–1245.
- International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921.
- Jordan, I.K., Rogozin, I.B., Glazko, G.V., and Koonin, E.V. (2003). Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends Genet.* **19**, 68–72.
- Jurka, J. (2000). Repbase update: a database and an electronic journal of repetitive elements. *Trends Genet.* **16**, 418–420.
- Keshet, E., Schiff, R., and Itin, A. (1991). Mouse retrotransposons: a cellular reservoir of long terminal repeat (LTR) elements with diverse transcriptional specificities. *Adv. Cancer Res.* **56**, 215–251.
- Kigami, D., Minami, N., Takayama, H., and Imai, H. (2003). MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos. *Biol. Reprod.* **68**, 651–654.
- Kondo, Y., and Issa, J.P. (2003). Enrichment for histone H3 lysine 9 methylation at Alu repeats in human cells. *J. Biol. Chem.* **278**, 27658–27662.
- Latham, K.E., and Schultz, R.M. (2001). Embryonic genome activation. *Front. Biosci.* **6**, D748–D759.
- Lawrence, C.E., Altschul, S.F., Boguski, M.S., Liu, J.S., Neuwald, A.F., and Wootton, J.C. (1993). Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. *Science* **262**, 208–214.
- Loftus, S.K., Morris, J.A., Carstea, E.D., Gu, J.Z., Cummings, C., Brown, A., Ellison, J., Ohno, K., Rosenfeld, M.A., Tagle, D.A., et al. (1997). Murine model of Niemann-Pick C disease: mutation in a cholesterol homeostasis gene. *Science* **277**, 232–235.
- Mayer, W., Niveleau, A., Walter, J., Fundele, R., and Haaf, T. (2000). Demethylation of the zygotic paternal genome. *Nature* **403**, 501–502.
- McClintock, B. (1950). The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. USA* **36**, 344–355.
- McClintock, B. (1953). Induction of instability at selected loci in maize. *Genetics* **38**, 579–599.
- Medstrand, P., van de Lagemaat, L.N., and Mager, D.L. (2002). Retroelement distributions in the human genome: variations associated with age and proximity to genes. *Genome Res.* **12**, 1483–1495.
- Miller, I., and Freund, J.E. (1965). *Probability and Statistics for Engineers* (Englewood Cliffs, NJ: Prentice-Hall, Inc.).
- Miller, W.J., McDonald, J.F., Nouaud, D., and Anxolabehere, D. (1999). Molecular domestication—more than a sporadic episode in evolution. *Genetica* **107**, 197–207.
- Monk, M., Boubelik, M., and Lehnert, S. (1987). Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development* **99**, 371–382.
- Mouse Genome Sequencing Consortium. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562.
- Ning, Z., Cox, A.J., and Mullikin, J.C. (2001). SSAHA: a fast search method for large DNA databases. *Genome Res.* **11**, 1725–1729.
- Norton, J.D., and Hogan, B.L. (1988). Temporal and tissue-specific expression of distinct retrovirus-like (VL30) elements during mouse development. *Dev. Biol.* **125**, 226–228.
- Oh, B., Hwang, S.Y., Solter, D., and Knowles, B.B. (1997). Spindlin, a major maternal transcript expressed in the mouse during the transition from oocyte to embryo. *Development* **124**, 493–503.
- Oh, B., Hwang, S.-Y., de Vries, W.N., Solter, D., and Knowles, B.B. (1999). Identification of genes and processes guiding the transition between the mammalian gamete and embryo. In *Advances in Molecular Biology: A Comparative Methods Approach to the Study of Oocytes and Embryos*, J.D. Richter, ed. (Oxford: Oxford University Press), pp. 101–126.
- Oh, B., Hwang, S.-Y., McLaughlin, J., Solter, D., and Knowles, B.B. (2000). Timely translation during the mouse oocyte-to-embryo transition. *Development* **127**, 3795–3803.
- Orgel, L.E., and Crick, F.H. (1980). Selfish DNA: the ultimate parasite. *Nature* **284**, 604–607.
- Ostertag, E.M., DeBerardinis, R.J., Goodier, J.L., Zhang, Y., Yang, N., Gerton, G.L., and Kazazian, H.H., Jr. (2002). A mouse model of human L1 retrotransposition. *Nat. Genet.* **32**, 655–660.
- Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., Dean, W., Reik, W., and Walter, J. (2000). Active demethylation of the paternal genome in the mouse zygote. *Curr. Biol.* **10**, 475–478.
- Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S., and Sharp, P.A. (1986). Splicing of messenger RNA precursors. *Annu. Rev. Biochem.* **55**, 1119–1150.

- Piko, L., Hammons, M.D., and Taylor, K.D. (1984). Amounts, synthesis, and some properties of intracisternal A particle-related RNA in early mouse embryos. *Proc. Natl. Acad. Sci. USA* *81*, 488–492.
- Pittoggi, C., Sciamanna, I., Mattei, E., Beraldi, R., Lobascio, A.M., Mai, A., Quaglia, M.G., Lorenzini, R., and Spadafora, C. (2003). Role of endogenous reverse transcriptase in murine early embryo development. *Mol. Reprod. Dev.* *66*, 225–236.
- Poznanski, A.A., and Calarco, P.G. (1991). The expression of intracisternal A particle genes in the preimplantation mouse embryo. *Dev. Biol.* *143*, 271–281.
- Reese, M.G., Eeckman, F.H., Kulp, D., and Haussler, D. (1997). Improved splice site detection in Genie. *J. Comput. Biol.* *4*, 311–323.
- Rothstein, J.L., Johnson, D., DeLoia, J.A., Skowronski, J., Solter, D., and Knowles, B. (1992). Gene expression during preimplantation mouse development. *Genes Dev.* *6*, 1190–1201.
- Santos, F., Hendrich, B., Reik, W., and Dean, W. (2002). Dynamic reprogramming of DNA methylation in the early mouse embryo. *Dev. Biol.* *241*, 172–182.
- Schramke, V., and Allshire, R. (2003). Hairpin RNAs and retrotransposon LTRs effect RNAi and chromatin-based gene silencing. *Science* *301*, 1069–1074.
- Smit, A.F. (1993). Identification of a new, abundant superfamily of mammalian LTR-transposons. *Nucleic Acids Res.* *21*, 1863–1872.
- Smit, A.F.A. (1996). Structure and evolution of mammalian interspersed repeats. PhD thesis, University of Southern California, Los Angeles, Los Angeles, California.
- Smit, A.F. (1999). Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.* *9*, 657–663.
- Solter, D., and Knowles, B.B. (1975). Immunosurgery of mouse blastocyst. *Proc. Natl. Acad. Sci. USA* *72*, 5099–5102.
- Speek, M. (2001). Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol. Cell. Biol.* *21*, 1973–1985.
- Staley, J.P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* *92*, 315–326.
- Svoboda, P., Stein, P., Hayashi, H., and Schultz, R.M. (2000). Selective reduction of dormant maternal mRNAs in mouse oocytes by RNA interference. *Development* *127*, 4147–4156.
- Svoboda, P., Stein, P., Anger, M., Bernstein, E., Hannon, G.J., and Schultz, R.M. (2004). RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. *Dev. Biol.* *269*, 276–285.
- Taylor, K.D., and Piko, L. (1987). Patterns of mRNA prevalence and expression of B1 and B2 transcripts in early mouse embryos. *Development* *101*, 877–892.
- van de Lagemaat, L.N., Landry, J.-R., Mager, D.L., and Medstrand, P. (2003). Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends Genet.* *90*, 530–536.
- Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I., and Martienssen, R.A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* *297*, 1833–1837.
- Walsh, C.P., and Bestor, T.H. (1999). Cytosine methylation and mammalian development. *Genes Dev.* *13*, 26–34.
- Wang, Q.T., Piotrowska, K., Ciemerych, M.A., Milenkovic, L., Scott, M.P., Davis, R.W., and Zernicka-Goetz, M. (2004). A genome-wide study of gene activity reveals developmental signaling pathways in the preimplantation mouse embryo. *Dev. Cell* *6*, 133–144.
- Yoder, J.A., Walsh, C.P., and Bestor, T.H. (1997). Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* *13*, 335–340.