Genome plasticity in the mouse oocyte and early embryo

A.E. Peaston*, B.B. Knowles and K.W. Hutchison†
†Department of Biochemistry, Microbiology and Molecular Biology, The University of Maine, Orono, ME 04469, U.S.A.

Abstract
In dissecting the molecules and molecular mechanisms that control mammalian oocyte-to-embryo transition, we found abundant transcripts representing developmentally regulated ERVs (endogenous retroviruses) in mouse oocyte and two-cell stage embryo cDNA libraries. These retrotransposons can act as alternative promoters and first exons for diverse genes, synchronizing their expression. Heritable genetic change due to replication of these retrotransposons probably occurs specifically in oocytes and early embryos. ERVs are usually epigenetically silenced, through DNA methylation and chromatin-based mechanisms. Their activation and silencing indicates a change in the epigenetic state of the genome. The thousands of endogenous retro-elements in the mouse genome provides potential scope for large-scale co-ordinated epigenetic fluctuations and leads to the hypothesis that differential transposable element expression triggers sequential reprogramming of the embryonic genome during the oocyte-to-embryo transition.

Introduction
Genome plasticity refers to the propensity for the genome to undergo heritable permanent change. In metazoans, in addition to changes brought about specifically by meiotic recombination in sexual reproduction, variable size DNA deletions, insertions, duplications, inversions, translocations and single nucleotide mutations occur. To be heritable, these changes must occur in germ cells, gametes or cells of the early embryo destined to contribute to the germ cell lineage. Do cellular processes that affect the genome specifically at these times affect the type of genomic alteration likely to occur at each stage? Emerging understanding of mouse oocyte and early embryo molecular physiology provides evidence of past changes with specific impact on oocytes, as well as clues to the potential agents of ongoing plasticity.

The success of early development in sexually reproducing metazoans depends on the union of the genomes of the oocyte and sperm to form the genome of the totipotent embryo. Molecules necessary for this transition are laid down in the growing oocyte. After transcription becomes undetectable in the FGO (full-grown oocyte), the completion of meiotic maturation, fertilization, gametic genome reprogramming and activation of the new embryonic genome are completely dependent on timely controlled translation of the stored maternal RNAs and degradation of proteins and mRNAs no longer required. To establish the molecular basis of the transition from oocyte to embryo in mice, we randomly sampled, sequenced and analysed large representative cDNA libraries prepared from FGO and two-cell stage embryos [1–3]. Recent comparison of these libraries with each other, and comparison of the FGO library with equivalent oocyte libraries from two other chordates, allowed inference of conserved genes and molecular, biochemical and cellular processes underlying the initiation of chordate development [3]. Unique features of the FGO and two-cell stage embryo libraries shed light on genome plasticity potentially originating in these stages.

Analysis of the FGO library showed expression of approx. 4800 genes, approx. 10% of which are expressed only in oocytes [3]. Sixty highly expressed FGO-restricted genes fall into six families of clustered tandemly duplicated paralogues, Fbxw, Nalp, Obox, Oog, Oas1 and Tcl1b [3]. Another mechanism of gene duplication is reverse transcription of mRNA and random reintegration of the resulting intronless cDNA. Eight of 14 ‘retrogenes’ transcribed in the FGO are expressed exclusively in oocytes and lack human orthologues, indicating retrotransposition after the human/mouse split [3]. Thus both segmental errors of recombination or replication and retrotransposition have been important sources of new material for mammalian species-specific oocyte transcripts.

Retrotransposition expression, predominantly of LTR (long terminal repeat) Class III elements resembling ERVs (endogenous retroviruses), accounts for 14% and approx. 3% of the ESTs (expressed sequence tags) in the FGO and two-cell stage cDNA libraries respectively. Expression of these and other retrotransposons is developmentally regulated, with the MTA [MT (mouse transcript retrotransposon) subfamily A] [4,5] dominating in the oocyte, and MuERV-L (murine ERV-L) [6] dominating in the two-cell stage embryo [1,3]. ESTs of these two and other phylogenetically young

©2007 Biochemical Society
ERVs in the FGO and two-cell stage cDNA libraries outnumber ESTs of SINEs and LINEs (short and long interspersed nuclear elements respectively) whose genomic copy numbers are far greater than the ERVs.

MT elements are particularly abundantly expressed in the oocyte, under strict developmental regulation. Genomic evidence for their recent transpositional activity suggests that they may influence continuing evolution of the mouse genome [5]. MT elements have a basic ERV structure with approx. 396 bp LTRs flanking an internal sequence. However, the internal sequence lacks recognizable ORFs (open reading frames), and thus MTs are transpositionally non-autonomous [5]. MuERV-L is currently active in the mouse genome and contains an uninterrupted ORF encoding Gag and Pol [6,7]. MT and MuERV-L elements share flanking 5 bp target site duplications (A.E. Peaston, unpublished work, and [8]), and have overlapping expression patterns, suggesting that MuERV-L may be responsible for the mobilization in trans of MT. In the present study, we show results demonstrating that newly integrated MTs are accumulating in C57BL/6J mice. Our results suggest that these full-length elements are more likely to be expressed in the FGO than older full-length elements and that expression is unlikely to be directly linked with CpG methylation of MT LTRs.

Methods
To estimate the MTA content of the mouse genome, coordinates for all MTA LTRs and MT-ints (internal sequences of MT) were downloaded from the UCSC Genome Browser (http://genome.ucsc.edu/) and computationally divided into: (i) solo LTRs and (ii) putative intact MTA elements defined as two MTA LTRs in the same orientation and separated by an internal sequence of ≤1.2 kb. To identify expressed intact MTAs, FGO ESTs identified as MTA and MT-int were aligned to the mouse genome using BLAT and collapsed into contigs of high stringency (100% identity). A bona fide expressed full-length MTA was defined as a contig extending from the 5′-LTR to the 3′-LTR of one or more intact MTAs in the genome.

To determine the strain-specific presence or absence of a putatively expressed MTA located at Chr2:38524664 (Mouse Genome Build 33), genomic DNA from four different inbred mouse strains (C57BL/6j, 129/Sv, DBA/2J or A/J) was used as the template for PCR amplification. Primers were designed from the 5′ and 3′ regions flanking the full-length MTA in C57BL/6j.

To assess the methylation status of developmentally regulated LTRs, DNA was isolated from aliquots of 50 oocytes or embryos and bisulfite-treated using the MethylEasy™ kit (Human Genetic Signatures Pty Ltd). Locus-specific PCR was conducted on aliquots equivalent to one to five embryos, and the products were cloned using pGEM® T-Easy (Promega). Clones were sequenced and accepted for CpG analysis if >95% of the non-CpG cytosines were converted into thymidine.

Further details of primers, PCR materials and conditions are available on request.

Table 1 | Evidence that expressed MTA loci are recent integrations into the mouse genome

<table>
<thead>
<tr>
<th></th>
<th>Expressed</th>
<th></th>
<th>Non-expressed</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of loci</td>
<td>LTR identity (%)</td>
<td>No. of loci</td>
<td>LTR identity (%)</td>
</tr>
<tr>
<td>C57BL/6j unique</td>
<td>12</td>
<td>99.17</td>
<td>3</td>
<td>95.67</td>
</tr>
<tr>
<td>Multi-strain</td>
<td>12</td>
<td>99.75</td>
<td>23</td>
<td>95.78</td>
</tr>
<tr>
<td>Total</td>
<td>24</td>
<td></td>
<td>26</td>
<td></td>
</tr>
</tbody>
</table>

Results
We estimated that there are 9837 MTA LTRs in the mouse genome, divided into 7401 solo LTRs and 1218 full-length elements, distributed among all chromosomes. This estimate is broadly in line with previous estimates of total MT copy number and makes MTA one of the most abundant retro elements in the mouse genome [4,8]. This raises the question as to whether the high level of MTA transcripts in the FGO is the result of low-level expression of all or most MTA loci or high-level expression of relatively few. To address this question we aligned the MTA ESTs to the genome, looking for the ability to build a contig from the 5′-LTR to the 3′-LTR. Our analysis suggested that perhaps as few as 36 intact MTA loci are expressed in the FGO. This number is probably a conservative estimate of transcriptionally active MTA loci, since our analysis used high-stringency matching of the full-length elements and does not account for transcriptionally active solitary MTA LTRs. Furthermore, low-abundance MTA transcripts may not be represented in the FGO library and, consequently, will not be considered here.

Active expression may indicate loci whose activity is selected and maintained through mouse genome evolution. Alternatively, actively expressed loci could be the most recent MTA integrations into the mouse genome that have not yet been silenced either epigenetically or through mutation. To resolve this question, the level of sequence identity between the 5′- and 3′-LTR was compared for each putatively expressed MTA locus and for a set of MTA elements without evidence of expression in the FGO cDNA library. Upon integration of an ERV, the 5′- and 3′-LTRs become independent genetic entities, free to diverge from each other. As shown in Table 1, the putatively expressed MTA loci had a level of sequence identity greater than 99% between the 5′- and 3′-LTRs. Most of the differences were single base substitutions (results not shown). In contrast, the non-expressed loci had a lower level of sequence identity (<95%) between their 5′- and 3′-LTRs. Divergence included not only single base substitutions, but a greater number of insertions/deletions.
The predicted size of the C57BL/6J amplicon containing the MT element plus the flanking sequences is 2207 kb (+MT). The predicted size of the amplicon without an MT insertion is 0.325 kb (−MT). The mouse strain from which the template genomic DNA was isolated for PCR amplification is indicated at the top. Sizes of some bands of the DNA molecular-mass markers (MWM) are indicated on the right. NS indicates a non-specific amplicon present in all PCR reactions.

Expression of LTR elements and other retrotransposons has been reported for many normal tissues and cell types, but is thought to be usually repressed by a mechanism involving cytosine methylation of CpG dinucleotides together with repressive chromatin modifications to each locus. In general, cytosine methylation of CpG dinucleotides occurs somewhat unevenly between the maternal and paternal genomes, but genomewide demethylation begins in the two-cell stage embryo and remethylation occurs in the blastocyst and after implantation [9]. Studies with different retrotransposable elements suggest also that retrotransposons at individual loci may vary considerably from the general case [10,11]. These studies examined elements with a relatively high density of CpG dinucleotides in their promoter or LTR modules.

Analysis of MTAs and MTBs (MT subfamily B) driving chimaeric gene expression in the oocyte showed, on average, less than one CpG site per LTR (results not shown). We analysed CpG methylation in the solo LTRs driving the expression of Spin, Nfil3 and Zbed3. Transcription of these genes in oocytes and early embryos is driven by MTB (Spin) and MTC (MT subfamily C) (Nfil3), and by the MT2B (Zbed3) element respectively. MTB and MTC are members of the MT family, whereas MT2B is thought to be a more distant relative of the MT family [8]. These genes were chosen since unequivocal genomic loci of the LTRs allowed their specific and detailed analysis using bisulfite sequencing, and we had previously established their expression patterns in oocytes and pre-implantation embryos [2]. The LTRs examined are between 396 and 600 nt long and contain one CpG (Spin – MTB), three CpG (Nfil3 – MTC) and seven CpG (Zbed3 – MT2B) sites. The MT2B element driving expression of Zbed3 is completely unmethylated in the morula and blastocyst (Figure 2), at which time Zbed3 expression is silenced or greatly reduced from earlier stages [2]. The single CpG of the MTB element driving expression of Spin is methylated in oocytes when transcripts are abundant, but substantially unmethylated from the zygote to four-cell stage embryos as transcripts decline to an undetectable level (Figure 2) [2]. Notably, the single CpG in this LTR is methylated well in advance of the methylation of a CpG in the intronic sequence 16 nt 5' of the LTR. This suggests that the silencing machinery distinguishes the LTR from flanking intronic sequence. Finally, Nfil3 transcription from MTC is silenced after the early two-cell stage, yet the MTC LTR remains unmethylated throughout the oocyte-to-embryo transition and at least to the blastocyst stage (Figure 2) [2]. The apparent transient zygotic increase in methylation of the LTRs driving Zbed3 and Spin probably reflects methylated allele contribution from the paternal genome, which is demethylated by the two-cell stage.
PCR-generated clones of bisulfite-treated DNA were sequenced to determine the methylation status of all CpG dinucleotides in the LTRs. Circles represent a single clone from a single PCR, including the target LTR and a small amount of flanking sequence. Each stage shows results from two to five independent sample preparations. Circles represent individual CpG dinucleotides, closed circles represent CpGs with a methylated cytosine, and open circles represent CpGs with an unmethylated cytosine. Absent circles indicate a clone that was incompletely sequenced. The position of the circles shows the approximate CpG positions in the LTR. The bar labelled 5′-LTR indicates the Spin LTR, and the circle upstream is a CpG in the upstream flanking DNA. Ovo, ovulated oocyte; Zyg, zygote; L2C, late two-cell embryo; 4-8C, four- to eight-cell stage embryo; Mor, morula; Blast, blastocyst; Somatic, DNA derived from mouse tail or spleen.

**Discussion**

Our results indicating multiple allelic differences for MTA insertions between different strains of inbred mice provide concrete evidence of considerable current MTA retrotranspositional activity. This activity is likely to occur in oocytes or two-cell stage embryos since that is when MTA transcripts are highly abundant, potentially making MTA-derived mouse genome plasticity a maternally regulated effect. New genomic retrotransposon insertions can cause coding sequence or regulatory mutations to normal cellular genes. Multiple insertions of identical controlling regions of transposable elements has been suggested as one way by which, over evolutionary time spans, new co-ordinately regulated gene networks might arise [12]. Indeed, one peculiarity of the FGO transcriptome is its high content of co-ordinately regulated chimaeric transcripts with alternative promoter and first exon provided by LTRs, many from MTs [2].

Conservation of synteny with rat for some of these suggests evolutionary selection and gain of function of the regulatory unit in the host [2]. A single new MT insertional mutation has been recorded in which the insertion caused a frameshift mutation resulting in a truncated protein [13].

Evidence suggests that enrichment for repetitive elements in the genome pre-disposes homology-driven unequal recombination and may foster the generation of segmental DNA duplications [14–16]. Indeed, MT loci were noted previously as recombination hotspots ([8] and references therein). An estimated two-thirds of the mouse and human genomes are duplicates [17,18]. Gene duplication generates functional redundancy, and, over evolutionary time, the potential for loss or gain of function and/or expression. This is exemplified in the six families of tandem duplicates mentioned in the Introduction. Despite their genomic maintenance in clusters and oocyte-specific expression, data from targeted deletion studies suggest that the paralogues are not functionally redundant and have been subject to purifying selection [3,19,20].

Functional retrotransposon proteins potentially predispose normal cellular gene transcripts to duplication by retrotransposition. We determined that the activity of RT (reverse transcriptase) is two to three orders of magnitude greater in oocytes and pre-implantation embryos than somatic cells [1]. Whether this activity is due to ERV RT or RT from other sources with lower abundance transcripts such as LINEs is unknown. Similarly, it is unknown whether and what functional endonuclease and integrase activities are contained in oocytes and early embryos. Direct measurement of the rate of retrotransposition of de novo reverse-transcribed genes in oocytes and early embryos has not been reported. However, genes expressed in the two-cell library were more likely to have retrotransposed duplicates when compared with somatically expressed genes not found in oocyte, ovary, testis or pre-implantation embryo libraries [1]. Most processed pseudogenes are thought to be generated by LINEs [21]. RT-encoding transcripts in mouse oocytes and early embryos are overwhelmingly ERVs, raising the possibility that their translated products actively generate processed pseudogenes at this time in development. Together, these results suggest that the oocyte and two-cell stage embryo are likely to be rich sources of new retrotransposed material in mouse and perhaps other mammalian genomes.

Allelic differences for MTA insertions between different strains raises the question of whether the remarkably high level of MTA expression in the FGO is specific to the C57BL/6J strain. It is recognized that the probability of insertion of some ERVs is influenced by both strain background and ERV subtype, perhaps reflecting strain differences in the silencing machinery (reviewed in [22]). Investigation of MT expression in oocytes and early embryos of different strains raises the question of whether the remarkably high level of MTA expression in the FGO is specific to the C57BL/6J strain. It is recognized that the probability of insertion of some ERVs is influenced by both strain background and ERV subtype, perhaps reflecting strain differences in the silencing machinery (reviewed in [22]). Investigation of MT expression in oocytes and early embryos of different strains raises the question of whether the remarkably high level of MTA expression in the FGO is specific to the C57BL/6J strain. It is recognized that the probability of insertion of some ERVs is influenced by both strain background and ERV subtype, perhaps reflecting strain differences in the silencing machinery (reviewed in [22]). Investigation of MT expression in oocytes and early embryos of different strains raises the question of whether the remarkably high level of MTA expression in the FGO is specific to the C57BL/6J strain. It is recognized that the probability of insertion of some ERVs is influenced by both strain background and ERV subtype, perhaps reflecting strain differences in the silencing machinery (reviewed in [22]). Investigation of MT expression in oocytes and early embryos of different strains raises the question of whether the remarkably high level of MTA expression in the FGO is specific to the C57BL/6J strain. It is recognized that the probability of insertion of some ERVs is influenced by both strain background and ERV subtype, perhaps reflecting strain differences in the silencing machinery (reviewed in [22]). Investigation of MT expression in oocytes and early embryos of different strains raises the question of whether the remarkably high level of MTA expression in the FGO is specific to the C57BL/6J strain. It is recognized that the probability of insertion of some ERVs is influenced by both strain background and ERV subtype, perhaps reflecting strain differences in the silencing machinery (reviewed in [22]). Investigation of MT expression in oocytes and early embryos of different strains raises the question of whether the remarkably high level of MTA expression in the FGO is specific to the C57BL/6J strain. It is recognized that the probability of insertion of some ERVs is influenced by both strain background and ERV subtype, perhaps reflecting strain differences in the silencing machinery (reviewed in [22]). Investigation of MT expression in oocytes and early embryos of different strains raises the question of whether the remarkably high level of MTA expression in the FGO is specific to the C57BL/6J strain. It is recognized that the probability of insertion of some ERVs is influenced by both strain background and ERV subtype, perhaps reflecting strain differences in the silencing machinery (reviewed in [22]). Investigation of MT expression in oocytes and early embryos of different strains raises the question of whether the remarkably high level of MTA expression in the FGO is specific to the C57BL/6J strain. It is recognized that the probability of insertion of some ERVs is influenced by both strain background and ERV subtype, perhaps reflecting strain differences in the silencing machinery (reviewed in [22]). Investigation of MT expression in oocytes and early embryos of different strains raises the question of whether the remarkably high level of MTA expression in the FGO is specific to the C57BL/6J strain. It is recognized that the probability of insertion of some ERVs is influenced by both strain background and ERV subtype, perhaps reflecting strain differences in the silencing machinery (reviewed in [22]). Investigation of MT expression in oocytes and early embryos of different strains raises the question of whether the remarkably high level of MTA expression in the FGO is specific to the C57BL/6J strain. It is recognized that the probability of insertion of some ERVs is influenced by both strain background and ERV subtype, perhaps reflecting strain differences in the silencing machinery (reviewed in [22]). Investigation of MT expression in oocytes and early embryos of different strains raises the question of whether the remarkably high level of MTA expression in the FGO is specific to the C57BL/6J strain. It is recognized that the probability of insertion of some ERVs is influenced by both strain background and ERV subtype, perhaps reflecting strain differences in the silencing machinery (reviewed in [22]). Investigation of MT expression in oocytes and early embryos of different strains raises the question of whether the remarkably high level of MTA expression in the FGO is specific to the C57BL/6J strain. It is recognized that the probability of insertion of some ERVs is influenced by both strain background and ERV subtype, perhaps reflecting strain differences in the silencing machinery (reviewed in [22]).
of only a few CpG dinucleotides at any locus might be insufficient to attract and firmly hold the methyl-binding proteins and other chromatin-modifying proteins necessary for transcriptional silencing. However, this cannot be the explanation for silencing of the particular elements in the absence of CpG methylation. It is possible that non-CpG cytosine methylation contributes to silencing in these elements, as has been suggested by others in a different context [23]. An alternative explanation is that other chromatin-based modifications, together with possible fluctuations of appropriate transcription factors, might explain the pattern of gene silencing for these elements. Others have shown for HERV-K (human endogenous retrovirus K) that CpG demethylation will not activate transcription in the absence of the appropriate mix of transcription factors [24].

Our results suggest that recent MTA integrations are more likely to be expressed than older integrations. Further experiments using RT–PCR and RNA FISH (fluorescence in situ hybridization) are planned to confirm the findings and to experimentally gauge the genomic extent of MT transcriptional activation in the oocyte-to-embryo transition. We have as yet found no outstanding features of MTA LTRs that might explain the apparent difference in expression patterns between older and younger MTA LTRs (K.W. Hutchison, unpublished work). Chromatin factors, perhaps in combination with position effects, might be an important key in regulating MT activation.

Chromatin-based regulation of this multicycle dispersed repetitive element may provide a unique view of genome reprogramming at the outset of development.

We thank LanYing Shi and Andrea Holbrook for technical assistance. For critical reading and discussion we are grateful to Joel Graber, Mimi de Vries, Karen Fancher and Alexei Evsikov. This study was supported by grants from the USPHS (United States Public Health Service) NIH (National Institutes of Health; R01HD037102, F33HD049264 and P40CA34196) and the National Science Foundation (EJPS-0132384).

References


Received 13 November 2006