Retrogenes in Preimplantation Embryo Development: A Unique Mode of Transcriptional Regulation

Chiu-Jung Huang[1], Kong-Bung Choo[2,3,4]*

[1]*Department of Animal Science and Institute of Biotechnology, School of Agriculture, Chinese Culture University,*
[2]*Recombinant DNA Laboratory, Department of Medical Research and Education, Taipei Veterans General Hospital,*
[3]*Department of Biotechnology and Laboratory Science in Medicine, National Yang-Ming University, and*
[4]*Institute of Clinical Medicine, National Yang-Ming University School of Medicine, Taipei, Taiwan, R.O.C.*

Our studies show that retrogenes are preferentially expressed in preimplantation embryos. These genes carry a short noncoding exon 1 that contributes directly to expression of the gene, and a second exon that contains the coding sequence without intron interruption. We show that preimplantation gene expression is first regulated by developmentally regulated transcription factors that target exon 1 and the solitary intron, followed by promoter hypermethylation on implantation and in adult tissues. An understanding of the mechanisms of gene expression during preimplantation development should have an impact on the understanding and treatment of spontaneous abortion and infertility. [*J Chin Med Assoc* 2009;72(7):346–350]

Key Words: embryo development, methylation, preimplantation, retrogene

Preimplantation Development: Decision on To Be or Not To Be

The development of a mammalian embryo really begins at the genesis and maturation of the germ cells, namely, the egg and the sperm. When an egg is fertilized by a sperm, and when the 2 parental nuclei meet and fuse, a 1-cell zygote is formed. The 2 sets of haploidic parental chromosomes are now diploid and ready to develop into a unique and independent life form. This newly constituted very early embryonic entity now begins to undergo a small number of cycles of cleavage while journeying through the oviduct. When the embryo finally develops into a blastocyst at the end of the journey through the oviduct, it enters the uterus and is ready for implantation at an appropriate site on the uterine wall, marking an end to the decisive preimplantation development.

Understanding the molecular mechanisms that regulate gene expression during preimplantation-stage development is important for a number of obvious clinical reasons. Firstly, light would be shed on infertility, miscarriages and other cases of failed pregnancy. An understanding of preimplantation gene expression would possibly lead to contraceptive strategies that are safer and less interruptive, and also to treatment of infertility. Furthermore, the final stage of preimplantation development generates a blastocyst. It is from the inner cell mass of the blastocyst that the totipotent embryonic stem cell lines are established which bear the promise of being exploited in regenerative medicine barring legal and moral issues. What may not be too obvious is that since preimplantation development involves active, accurate and timely differentiation and cell cleavage, understanding such events at the molecular level may provide clues on how dysregulation of

*Correspondence to: Dr Kong-Bung Choo, Department of Medical Research and Education, Taipei Veterans General Hospital, 201, Section 2, Shih-Pai Road, Taipei 112, Taiwan, R.O.C.
E-mail: kbchu@vghtpe.gov.tw • Received: October 31, 2008 • Accepted: April 27, 2009

© 2009 Elsevier. All rights reserved.

these processes may lead to cancer. Oncogenes and tumor suppressor genes have, indeed, been regularly shown to be indispensable developmental genes, and vice versa.[1–4]

In our social structure, mothers play a crucial role in the nursing of the young. Likewise, on fertilization, the sperm contributes the paternal set of chromosomes and does little else. It is up to the maternal care of the proteins in the cytoplasm of the egg to nurse the newly constituted zygote in the very first stage of development. To allow independent development of the embryo, maternal gene expression is immediately shut down on successful fertilization and maternal transcripts are rapidly degraded; by the 2-cell stage, >90% of maternal mRNA is gone.[5,6] The embryo has to start to learn to stand on its own: the zygotic genome now needs to be turned on appropriately by the process of active chromatin restructuring.[7,8] Zygotic genome activation is initiated at the late 1-cell stage and is in full swing at the 2-cell stage, or else the embryo would fail to survive and be aborted. Studies have shown that this very first attempt of the zygote to independently express its own genome is rather unsteady and haphazard. This is reflected in the observation that expression of the zygotic genome is rather careless and promiscuous when apparently functionless genomic sequences such as the highly repetitive sequences are also transcribed.[9] Or, is the transcription of such sequences really meaningless? No one really knows the answers because few laboratories work on such issues.

Zygotic Genome Activation: First Round in Gene Expression

When a baby learns to walk, the initial steps are inevitably tentative, but wobbly steps soon turn into steady strides. At the end of the 2-cell stage of preimplantation embryo development, the full weight of accurate regulatory censorship is now in place, or the embryo fails. An important cellular first process for the burgeoning embryo is the first cell cleavage which is preceded by the first S-phase of the mitotic cell cycle involving the very first round of accurate DNA replication. Recombination between the maternal and paternal chromosomes is also an inevitable event to eventually generate a unique individual, albeit one bearing resemblances to the parents. Meanwhile, transcription of genes essential for preimplantation development swings into full throttle despite some degree of inaccuracies in the very first round of transcription.

Does the developing preimplantation embryo follow the full set of rules of gene expression as in somatic cells? Few laboratories have focused on this issue mainly because of scarcity of study materials and technical difficulties. For obvious moral and legal reasons, human preimplantation embryos are not readily available. Primate embryos are equally hard to come by. To study preimplantation embryo development, scientists inevitably turn to the use of rodents, a mammal closest in gene constitution to the human. Elucidation of the mode of gene expression in the preimplantation embryo is vastly sped up by the complete or draft versions of the genomes of many model organisms including the mouse and the rat. Free access to the completed human genome sequence permits researchers to rapidly assess and evaluate the possible clinical significance of their findings based on the convenient use of other model organisms.

Preimplantation Gene Expression: Regular Rules Do Not Always Apply

By a combination of bioinformatic search and laboratory evaluation, our laboratory first derived 3 genes, viz. *Rnf33* (previously called *2czf45*, NCBI GenBank accession number AF290197), *Zfp352* (*2czf48*, AF290196) and *Tdpoz1* (*2czf56*, AF290198), that are first expressed in the egg of the mouse.[10] The former 2 genes were members of the zinc-finger protein superfamily, and *Tdpoz1* was later found to be a member of a newly defined protein family designated by our laboratory as the TDPOZ family.[11] Despite probable exhaustive degradation of the maternal transcripts of these genes on fertilization, transcription of these genes is reactivated using the zygotic genome as the blueprint, and the zygotic transcripts are maintained up to at least the 8-cell stage; the transcripts are undetectable in the blastocyst prior to implantation and in later developmental stages, and in all the tissues and organs of the adult mouse tested.[10] However, when a more sensitive reverse transcription–PCR protocol of recent development was used in re-examination of expression of these genes in the adult organs, we were able to detect expression of some of these genes in the ovary and/or testis (Huang et al, manuscript submitted; also unpublished data). The detection in the ovary could be a reflection of expression in the egg. More recent reports have shown promiscuous gene expression in the testis[12] not unlike that which occurs in the very first round of transcription in late 1-cell to early 2-cell-stage development. Possible correlation between expression of these genes in the very early embryo and in the testis is presently under investigation in our laboratory.

On detailed analysis of the gene structure of the *Rnf33* gene, a paralog *Rnf35* (GenBank accession number AY063495) is found located within ~11.5 kb upstream of *Rnf33*.[13,14] Likewise, a paralog *Zfp353* (GenBank accession number AF358728) of *Zfp352* is also found but is located on a different chromosome.[15] Most interestingly, bioinformatic searches using the *Tdpoz1* sequence as the query have led to the discovery of a great number of homologous sequences in species of both the animal and plant kingdoms, and we have called this new gene family TDPOZ, for proteins with the TRAF domain (TD) and POZ/BTB domains.[11] In the mouse genome, our first searches have identified 5 *Tdpoz1*-like gene copies;[11] in the rat genome, 297 copies are found distributed in 7 chromosomes despite a major clustering on a single locus on chromosome 2.[16] In short, the 3 preimplantation developmental genes that were originally identified in our laboratory have now turned into 3 groups of highly homologous genes.

When cDNA and genomic sequences of these genes are compared, we found that all the genes uncovered are retrogenes that do not bear any introns in their coding sequences,[11,13–16] very much dissimilar to the normal gene structure which includes numerous intron sequences to extend the size of the structural genes to tens or hundreds of kilobase pairs long. Retrogenes are derivatives of some ancestral genes that most likely carried intron sequences. On transcription of some of the ancestral genes, the introns were removed by splicing. The intron-less mature ancestral transcripts then underwent a reverse transcription event to generate complementary DNA (cDNA) copies that were also free of introns. Some of such cDNA copies were successful in randomly reintegrating into the host genome. As the host evolved, the integrants were duplicated and transposed to other chromosomal sites or to different chromosomes while mutations were constantly being introduced into the parental or daughter copies (Figure 1).[15]

A cDNA sequence does not carry any promoter sequence since a cDNA sequence is the consequence of gene expression driven by the promoter. If integration of ancestral cDNA copies was random, how is expression of such retrogenes reactivated? We found clues in the structure of our preimplantation-specific genes. Intron-less retrogenes find transcriptional reactivation if they are inserted downstream of some active promoter sequences.[11,13–16] For example, the *Rnf35* gene is positioned downstream of a short exon the sequence of which constitutes the 5′-untranslated region (5′-UTR) of the mature *Rnf35* mRNA.[13,14] There is a 3.56-kb solitary intron separating the noncoding 5′-UTR exon 1 and the coding exon 2 of *Rnf35*. Likewise, the *Rnf33* carries a major promoter upstream of a similar noncoding 5′-UTR exon 1 that is utilized in gene expression in the preimplantation embryo. More intriguingly, we also detected in the preimplantation embryo other minor *Rnf33* transcripts that harbor other noncoding exons 1 located further upstream of the major exon 1; and 1 of these minor exon 1 sequences is identical to exon 1 of *Rnf35*, suggesting cotranscription of the *Rnf35–Rnf33* genes. Cotranscription of geographically adjacent genes is a rare event in eukaryotic gene expression, 99% of which follows the 1 gene–1 primary transcript

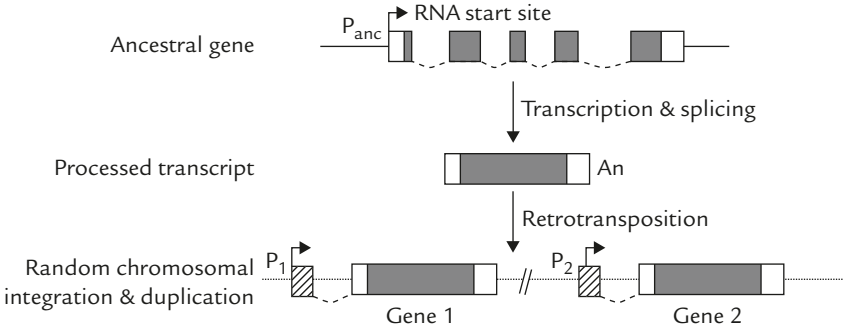

Figure 1. Derivation of retrogenes in the mammalian genome. In the simplistic model, the ancestral gene, which carried introns, was transcribed using an ancestral promoter (P_{anc}) and introns were removed by splicing. The processed transcript then underwent retrotransposition (a combination of reverse transcription and random chromosomal insertion) and chromosomal duplication to generate multiple gene copies, which further diversified through accumulation of mutational changes. Gene copies (e.g. Genes 1 & 2 in the figure) that have been transposed to chromosomal sites that harbor promoters and transcriptional regulatory sequences (e.g. P_1 and P_2) reacquire transcriptional activities and are transcribed. Such later-day genes would carry an intronless coding exon and 1 or more noncoding exons in the 5′-untranslated region of the transcripts. In the figure, the gray boxes represent coding sequences, the white boxes denote untranslated regions, and the hatched boxes are newly derived 5′-noncoding exons.

rule. However, such putative cotranscription of the *Rnf35–Rnf33* gene pairs occurs only during the initial phase of zygotic genome activation at the 2-cell stage, and there is no evidence of cotranscription beyond the 2-cell stage.

Do the 5′-UTR noncoding exon 1 and the solitary intron sequences contribute to the regulation of expression of the retrogenes under their care? Detailed analysis of *cis*-acting sequences in the upstream regulatory promoter, exon 1 and intron sequences has revealed that the preimplantation embryo-specific retrogenes do not normally use the ubiquitous TATA-box as a core promoter. For *Rnf35*, the core promoter is an initiator sequence that overlaps with the 5′-end of the exon 1, thus, assigning the first biological role to the exon 1 sequence.[17] Crucial protein factors identified that participate in *Rnf35* expression include the nuclear factor Y (NF-Y) and the repressor CCAAT-displacement protein (CDP).[17,18] Expression of the repressor *Cdp* gene first occurs at the 8-cell stage, coinciding with the shutting down of *Rnf35* expression. Subsequently, the *Rnf35* promoter sequence is hypermethylated (unpublished data) to relieve the CDP repression. Permanently freezing *Rnf35* gene expression by promoter hypermethylation is a mode of regulation that is economical in energy expenditure and frees the CDP protein for other biological functions.

Hence, our studies indicate 2 important aspects of preimplantation gene expression. There are important *cis*-acting sequences residing at exon 1 and the associated solitary intron. Furthermore, these *cis* sequences are targeted by developmentally regulated protein factors in stages: positive regulation by ubiquitous positive regulators such as NF-Y, followed by initial shutting down of gene expression by a repressor protein such as CDP (Figure 2). Subsequently, permanent gene silencing occurs via promoter hypermethylation. The importance of exon 1 and the solitary intron sequences in preimplantation gene expression is further supported in the analysis of the *Rnf33* gene, in which crucial positive and negative *cis*- and *trans*-acting elements have been identified in exon 1 and the solitary intron (Hsu, Huang and Choo, manuscript in preparation).

The Rat Rtdpoz Retrogenes: Developmental Regulation and Reactivation in Cancer Cells

When the sequence of the preimplantation-specific *Tdpoz1* mouse gene was used to query the rat genome sequence database, 297 hits were returned. A vast majority of the hits occurred within an ~1-Mb segment of the supercontig Rn2_047626 of chromosome 2.[16] Two of the hits represent the rat *Tdpoz* genes, termed *Rtdpoz-T1* and *Rtdpoz-T2*, which were shown to be expressed specifically in the rat testis. *T1* and *T2* are also retrogenes but carry 1 or more noncoding 5′-UTR exons in their transcripts (Choo et al, manuscript submitted). We showed that the *T1* and *T2* genes are developmentally regulated in that they are expressed until day 16 of development before the genes are silenced except for expression in the testis in adult rats. Surprisingly, *T1* and *T2* expression is detected in a number of rat cancer cell lines, including 2 pancreatic, a testicular cancer and an osteosarcoma cell line. Since besides the testis, we have strong evidence to indicate that normal bone and pancreas do not express *T1* and *T2*, our observation means there is reactivation of *T1* and *T2* in these cancer cells. We could not detect mutational changes in the promoter sequences

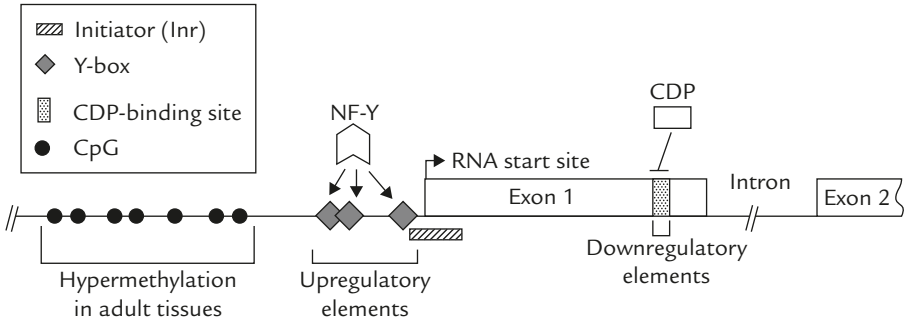

Figure 2. Transcriptional regulation of *Rnf35*, a model of preimplantation embryo-specific retrogene. The *Rnf35* gene carries a noncoding exon 1 and a solitary intron. Transcription of *Rnf35* is regulated by positive *cis*-acting elements including an initiator (Inr) core promoter (hatched box) and 3 Y-boxes (gray diamonds) that are targeted by the positive transcription factor, nuclear factor Y. An important negative *cis* element is found in exon 1 (dotted box) targeted by the negative regulator, CCAAT-displacement protein (CDP). CDP is first expressed at the 8-cell stage to exert the initial stage of *Rnf35* gene silencing. In the post-implantation embryo and in adult tissues, *Rnf35* is permanently silenced through methylation of CpG sites (denoted by black circles) located in the upstream regulatory sequence.

of these genes, and the promoter sequences of both normal and cancer cells were found to be hypomethylated, thus ruling out gene reactivation by DNA hypomethylation. The mechanism of gene reactivation of *T1* and *T2* in cancer cells is being investigated. Possible oncogenic contributions of the *T1* and *T2* proteins to the carcinogenesis process also remain to be elucidated.

Concluding Remarks

With the use of the mouse and rat model systems, our work has begun to shed light on the regulation of gene expression during the crucial preimplantation stage of embryo development. Although more genes need to be examined to get a truly representative picture, our seminal papers have revealed a rather unique gene structure for preimplantation-specific genes. These genes are all retrogenes in structure under the regulation of 1 or more noncoding 5′-UTR exons and associated intron sequences. Transcription factors that contribute to preimplantation gene regulation also appear to be developmentally regulated in accordance to the expression status of the preimplantation genes under regulation. On implantation, the genes are most likely silenced by hypermethylation. With such burgeoning understanding, we may now begin to ask if such a mode of regulation of gene expression is also strictly observed in embryos derived by animal cell cloning. Failure in continued development of early human embryo in miscarriage patients may also be examined on the basis of our findings with the hope of designing novel strategies to circumvent early embryo failures.

Acknowledgments

This work was supported by grants from the National Science Council (NSC94-2311-B-075-001, NSC95-2311-B-075-001, NSC96-2311-B-075-001) and Taipei Veterans General Hospital (V97C1-055 and V98C1-020), Taiwan, R.O.C.

References

1. Feldman BJ, Hampton T, Cleary ML. A carboxy-terminal deletion mutant of Notch1 accelerates lymphoid oncogenesis in E2A-PBX1 transgenic mice. *Blood* 2000;96:1906–13.
2. Lewis MT. Homeobox genes in mammary gland development and neoplasia. *Breast Cancer Res* 2000;2:158–69.
3. Thorsteinsdottir U, Kroon E, Jerome L, Blasi F, Sauvageau G. Defining roles for HOX and MEIS1 genes in induction of acute myeloid leukemia. *Mol Cell Biol* 2001;21:224–34.
4. Xie Y, Sun T, Wang QT, Wang Y, Wang F, Puscheck E, Rappolee DA. Acquisition of essential somatic cell cycle regulatory protein expression and implied activity occurs at the second to third cell division in mouse preimplantation embryos. *FEBS Lett* 2005;579:398–408.
5. Christians E, Campion E, Thompson EM, Renard JP. Expression of the HSP 70.1 gene, a landmark of early zygotic activity in the mouse embryo, is restricted to the first burst of transcription. *Development* 1995;121:113–22.
6. Aoki F, Worrad DM, Schultz RM. Regulation of transcriptional activity during the first and second cell cycles in the preimplantation mouse embryo. *Dev Biol* 1997;181:286–307.
7. Majumder S, Miranda M, DePamphilis ML. Analysis of gene expression in mouse preimplantation embryos demonstrates that the primary role of enhancers is to relieve repression of promoters. *EMBO J* 1993;12:1131–40.
8. Wiekowski M, Miranda M, DePamphilis ML. Requirements for promoter activity in mouse oocytes and embryos distinguish paternal pronuclei from maternal and zygotic nuclei. *Dev Biol* 1993;159:366–78.
9. Vinckenbosch N, Dupanloup I, Kaessmann H. Evolutionary fate of retroposed gene copies in the human genome. *Proc Natl Acad Sci USA* 2006;103:3220–5.
10. Choo KB, Chen HH, Cheng WT, Chang HS, Wang M. *In silico* mining of EST databases for novel pre-implantation embryo-specific zinc finger protein genes. *Mol Reprod Dev* 2001;59:249–55.
11. Huang CJ, Chen CY, Chen HH, Tsai SF, Choo KB. TDPOZ, a family of bipartite animal and plant proteins that contain the TRAF (TD) and POZ/BTB domains. *Gene* 2004;324:117–27.
12. Schmidt EE. Transcriptional promiscuity in testes. *Curr Biol* 1996;6:768–9.
13. Chen HH, Liu TY, Li H, Choo KB. Use of a common promoter by two juxtaposed and intronless mouse early embryonic genes, Rnf33 and Rnf35: implications in zygotic gene expression. *Genomics* 2002;80:140–3.
14. Choo KB, Chen HH, Liu TY, Chang CP. Different modes of regulation of transcription and pre-mRNA processing of the structurally juxtaposed homologs, Rnf33 and Rnf35, in eggs and in pre-implantation embryos. *Nucleic Acids Res* 2002;30:4836–44.
15. Chen HH, Liu TY, Huang CJ, Choo KB. Generation of two homologous and intronless zinc-finger protein genes, zfp352 and zfp353, with different expression patterns by retrotransposition. *Genomics* 2002;79:18–23.
16. Choo KB, Hsu MC, Chong KY, Huang CJ. Testis-specific expression and genomic multiplicity of the rat Rtdpoz genes that encode bipartite TRAF- and POZ/BTB-domain proteins. *Gene* 2007;387:141–9.
17. Huang CJ, Wu SC, Choo KB. Transcriptional modulation of the pre-implantation embryo-specific Rnf35 gene by the Y-box protein NF-Y/CBF. *Biochem J* 2005;387:367–75.
18. Huang CJ, Chang JG, Wu SC, Choo KB. Negative transcriptional modulation and silencing of the bi-exonic Rnf35 gene in the preimplantation embryo: binding of the CCAAT-displacement protein/Cux to the untranslated exon 1 sequence. *J Biol Chem* 2005;280:30681–8.