Identification of a Novel PNMA-MS1 Gene in Marsupials Suggests the LTR Retrotransposon-Derived PNMA Genes Evolved Differently in Marsupials and Eutherians

Sawa Iwasaki1,2, Shunsuke Suzuki3,4, Matthew Pelekanos3, Helen Clark3, Ryuichi Ono1, Geoff Shaw3, Marilyn B. Renfree3, Tomoko Kaneko-Ishino2, and Fumitoshi Ishino1,5,*

Department of Epigenetics, Medical Research Institute, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8510, Japan1; School of Health Sciences, Tokai University, Bohseidai, Isehara, Kanagawa 259-1193, Japan2; Department of Zoology, The University of Melbourne, Melbourne, Victoria 3010, Australia3; Epigenomics Division, Faculty of Agriculture, Frontier Agriscience and Technology Center, Shinshu University, 8304 Minami-minowa, Kami-ina, Nagano 399-4598, Japan4 and Global Center of Excellence Program for International Research Center for Molecular Science in Tooth and Bone Diseases, Tokyo Medical and Dental University, 1-5-4 S Yushima, Bunkyo-ku, Tokyo 113-8510, Japan5

*To whom correspondence should be addressed. Tel. +81-3-5803-4862. Fax. +81-3-5803-4863. Email: fishino.epgn@mri.tmd.ac.jp

Edited by Dr Toshihiko Shiroishi
(Received 5 November 2012; accepted 24 April 2013)

Abstract

Two major gene families derived from Ty3/Gypsy long terminal repeat (LTR) retrotransposons were recently identified in mammals. The sushi-ichi retrotransposon homologue (SIRH) family comprises 12 genes: 11 in eutherians including Peg10 and Peg11/Rtl1 that have essential roles in the eutherian placenta and 1 that is marsupial specific. Fifteen and 12 genes were reported in the second gene family, para-neoplastic antigen MA (PNMA), in humans and mice, respectively, although their biological functions and evolutionary history remain largely unknown. Here, we identified two novel candidate PNMA genes, PNMA-MS1 and -MS2 in marsupials. Like all eutherian-specific PNMA genes, they exhibit the highest homology to a Gypsy12_DR (DR, Danio rerio) Gag protein. PNMA-MS1 is conserved in both Australian and South American marsupial species, the tammar wallaby and grey short-tailed opossum. However, no PNMA-MS1 orthologue was found in eutherians, monotremes or non-mammalian vertebrates. PNMA-MS1 was expressed in the ovary, mammary gland and brain during development and growth in the tammar, suggesting that PNMA-MS1 may have acquired a marsupial-specific function. However, PNMA-MS2 seems to be a pseudogene. The absence of marsupial orthologues of eutherian PNMA genes suggests that the retrotransposition events of the Gypsy12_DR-related retrotransposons that gave rise to the PNMA family occurred after the divergence of marsupials and eutherians.

Key words: LTR retrotransposons; PNMA family; marsupial-specific genes; mammalian evolution

1. Introduction

Approximately 40–50% of the mammalian genome is derived from transposable elements, such as retrotransposons and DNA transposons. The Ty3/Gypsy long terminal repeat (LTR) retrotransposons have been detected in various eukaryotic organisms including fungi, plants, insects, tunicates and echinoderms as well as in several vertebrates, such as fish, amphibians and reptiles, but not in mammals and birds. However, discrete regions within these elements have acquired new functions as novel endogenous genes and are highly conserved in marsupials and eutherians. Two major gene families derived from LTR retrotransposons were recently identified in mammals. The sushi-ichi retrotransposon homologue (SIRH) family comprises 12 genes: 11 in eutherians including Peg10 and Peg11/Rtl1 that have essential roles in the eutherian placenta and 1 that is marsupial specific. Fifteen and 12 genes were reported in the second gene family, para-neoplastic antigen MA (PNMA), in humans and mice, respectively, although their biological functions and evolutionary history remain largely unknown. Here, we identified two novel candidate PNMA genes, PNMA-MS1 and -MS2 in marsupials. Like all eutherian-specific PNMA genes, they exhibit the highest homology to a Gypsy12_DR (DR, Danio rerio) Gag protein. PNMA-MS1 is conserved in both Australian and South American marsupial species, the tammar wallaby and grey short-tailed opossum. However, no PNMA-MS1 orthologue was found in eutherians, monotremes or non-mammalian vertebrates. PNMA-MS1 was expressed in the ovary, mammary gland and brain during development and growth in the tammar, suggesting that PNMA-MS1 may have acquired a marsupial-specific function. However, PNMA-MS2 seems to be a pseudogene. The absence of marsupial orthologues of eutherian PNMA genes suggests that the retrotransposition events of the Gypsy12_DR-related retrotransposons that gave rise to the PNMA family occurred after the divergence of marsupials and eutherians.
the Ty3/Gypsy LTR retrotransposons are the sushi-ichi retrotransposon homologue (SIRH) family (also called the MART or SUSHI family) comprising 12 genes encoding a Gag-like protein, each of which has 20–30% similarity to the sushi-ichi retrotransposon Gag in pufferfish (Takifugu rubripes),\(^9\)–\(^{11}\) and the para-neoplastic antigen MA (PNMA) family also encoding the Gag-like protein homologous to the Gypsy12_DR retrotransposon Gag in zebrafish (DR, Danio rerio)\(^{13}\) comprising 15 and 11 genes in humans and mice, respectively. It should be noted that the homology between these two LTR retrotransposons is only 6.5% and 13.6% along with the entire Gag and Pol regions, respectively. PEG10/SIRH1 is a therian-specific gene, and PEG11/SIRH2 and the remaining SIRH3–11 seem eutherian specific, while SIRH12 was derived from a marsupial-specific retrotransposition event. We previously demonstrated that Peg10/Sirh1 and Peg11/Sirh2 are essential for placentation and formation in mice.\(^7\),\(^8\) Most PNMA genes are expressed in the brains of macaques and mice and their functions remain unknown.\(^{14}\) PNMA1–3 were first identified as genes encoding neuronal auto-antigens using sera from patients with para-neoplastic neurological syndromes.\(^{15}\) Schüller et al.\(^{16}\) and Campillos et al.\(^{13}\) performed genome-wide analyses and identified additional 12 family genes in humans among which PNMA6 has no mouse orthologue. No Gypsy12_DR Gag-derived sequences were reported in birds,\(^{13}\) and thus, it is probable that the PNMA genes are also mammal specific. However, the search has been limited in several eutherian species and the existence of marsupial orthologues and/or marsupial-specific PNMA genes remained unknown.

Here, we conducted comprehensive in silico screening for the PNMA genes using the whole-genome shotgun (WGS) sequences of the grey short-tailed opossum\(^4\) and the tammar wallaby\(^5\) and identified a novel PNMA-MS1 gene as the first marsupial-specific PNMA gene.

2. Materials and methods

2.1. Animals and tissue collection

Tammar wallabies (Macropus eugenii) of Kangaroo Island, South Australia origin, were maintained in The University of Melbourne marsupial breeding colony in grassy outdoor enclosures. Lucerne cubes, grass and water were provided ad libitum and supplemented with fresh vegetables. The day of birth of pouch young was designated as d0. When the day of birth was unknown, their age was estimated using the head length.\(^{17}\) Foetal tissues including the head and body were collected from two foetuses at Day 23 and 26 of gestation, and the yolk sac placenta (YSP) from four foetuses sampled between Day 23 and 26 of gestation. Tissues including the brain, liver, lung, kidney, ovary and testis were collected from two pouched young aged Day 60–70 after birth. The liver, lung, pancreas, stomach, bladder, heart, kidney, adrenal, spleen and brain (cerebrum and cerebellum) were collected from Day 152 and 162 pouch young. Adult female tissues, including the brain (thalamus, hypothalamus and pituitary), ovary (ovary with active corpus luteum, corpus luteum, ovary with developing follicle and ovary with primary or secondary follicle), endometrium (gravid endometrium and non-gravid endometrium) and mammary gland (sucked gland and non-sucked gland) were also collected from two adults. Grey short-tailed opossums (Monodelphis domestica) were purchased from a breeding colony in the Department of Physiology at the University of Melbourne. The brain, liver, spleen, pituitary and ovary were collected from five adult opossums. Experimental procedures conformed to the Australian National Health and Medical Research Council guidelines\(^{18}\) and were approved by the Animal Experimentation Ethics Committees of the University of Melbourne.

2.2. Reverse transcriptase polymerase chain reaction

Genomic DNA and total RNA from tissues were prepared by using TRIZOL (Invitrogen), as described in the manufacturer’s protocol. cDNA was synthesized from 1 µg of total RNA using Superscript III reverse transcriptase (Invitrogen) with an oligo dT primer. Polymerase chain reaction (PCR) amplification for gene expression profiles were carried out using 10–100 ng of cDNA in a 25-µl reaction mixture containing 1 × ExTaq buffer, 2.5 mM deoxynucleotide triphosphates, 10 pmol primers and 1.25 U ExTaq HS (Takara) and were subjected to 30–35 PCR cycles; 96°C for 15 s, 60–65°C for 30 s and 72°C for 15–120 s depending on the length of PCR products at the ratio of 1 min/kb. PCR products were visualized by agarose gel electrophoresis with ethidium bromide staining. The primers used for the expression profiles were as follows: PNMAMS1-F1 (5'-AAC ATG GTG GAG TCT CTA GGA TGA TCC TG T G-3'), PNMAMS1-R1 (5'-CAA CGG TAA GGT GAC CTC TTA G G-3'), wGAPDH-F1 (5'-AGA AAG TGG TGA AGC AGG CAT-3'), wGAPDH-R1 (5'-TGG AGG ACA TGA AGG CCA TGA G-3'), wGAPDH-F2 (5'-CCT ACT CCA ATG TAT CTG TGT-3'), wGAPDH-R2 (5'-GTT GGA ACT CCT TTT TTG ACT G G-3'), LAMA3-F1 (5'-ACT CTA CCA AGA TCA GCA CAC C-3'), LAMA3-R1 (5'-CTG CCT CTA TCA AGA CAG T T-3'); PNMAMS2-F1 (5'-GCC TAA TGG AAA GTC ATA AGA AAG C-3'), PNMAMS2-R1 (5'-GAT TCC TGT ATA CAA ATG GTT GTC C-3'), PNMAMS2-F2 (5'-TTG ATG CAT TGT CTG AAA CCA G G-3'), PNMAMS2-R2 (5'-ATC TAT CAA CCA AGC GGC AAC T-3'); oOA21-F1 (5'-ATA AAC CCA GCA CCA CCG TCC ACG-3'), oOA21-R1 (5'-GTT CTC ACA ATC TCA AAG CCC AAA AAG-3').
2.3. 5′-and 3′-RACE

Rapid amplification of cDNA ends (RACE) reactions were performed with the tammar liver using the RNA SMARTER RACE cDNA Amplification kit (Clontech) according to the manufacturer’s recommendations. The 5′- and 3′-RACE fragments were generated with the following gene-specific primers: PNMAMS1-5′RACE-GSP1 (5′-TGC GTA TGG AGG GGA GAG TGA GCA AG-3′) and PNMAMS1-3′RACE-GSP1 (5′-GAC TGT GCC ATC GGG AGA AGG TGA AC-3′), and nested PCR was performed with following primers: PNMAMS1-5′RACE-GSP2 (5′-CAG ACA AGG TGG GGT 0 GCA AG-3′) and PNMAMS1-3′RACE-GSP2 (5′-TTT CTG TGA AGG TCT CCC TCT C-3′), respectively.

2.4. Detection and prediction of the Gypsy12_DR Gag-derived genes

For the detection of PNMA family genes, we performed TBLASTN searches (e-value <1.0E−9) using the NCBI server (http://blast.ncbi.nlm.nih.gov/Blast.cgi) against eutherian reference genomic sequences and marsupial WGS sequences using the Gypsy12_L1_Dr Gag protein sequence from Repbase (http://www.girinst.org/) as a query. After TBLASTN searches, sequences that encoded open reading frames (ORFs) with >100 aa (amino acids) were selected for the next analysis. Secondary screening was performed with all the sequences that were selected by first screening as a query. In addition, only sequences encoding proteins with >100 aa were considered to be candidate PNMA family genes and those with <100 aa were considered as PNMA pseudogenes. Genome resources used were: Homo sapiens (GRCh37.p5), Mus musculus (MGSCv37), M. eugenii (Meug_1.1) and M. domestica (MonDom5). ORF prediction was performed using an ORF finder (http://www.ncbi.nlm.nih.gov/orf/gorf.html).

2.5. Multiple alignment and phylogenetic tree

PNMA family genes and retrotransposon sequences from Repbase were aligned using the MEGA 5.0 (Molecular Evolutionary Genetics Analysis). Phylogenetic tree analysis was also performed using the MEGA 5.0. The tree was inferred using the neighbour-joining method with the bootstrap test (1000 replicates). The evolutionary distances were computed using the p-distance method and are in the units of the number of amino acid differences per site. All ambiguous positions were removed for each sequence pair.

2.6. Comparative genomic analysis

For comparison of the marsupial PNMA-MS1 and -MS2 genomic regions with the corresponding regions in eutherian species, we extracted the following sequences from Ensembl (http://www.ensembl.org);

PNMA-MS1; M. eugenii (GeneScaffold: Meug_1.0: 503: 24 720–39 177), M. domestica (Chromosome: MonDom5: 3: 260 858 116–260 881 268), H. sapiens (Chromosome: GRCh37: 18: 21 343 369–21 355 887), M. musculus (Chromosome: GRCh37: 18: 12 572 248–12 578 437), Ornithorhynchus anatinus (Chromosome: OANAS: 7: 17 676 195–17 684 806), Gallus gallus (Chromosome: WASHUC2: 2: 106 303 674–106 308 473), Xenopus tropicalis (Scaffold: JGI4.1: 84: 2861 611–2864 246), T. rubripes (Scaffold: FUGU4: 285: 98 142–99 106) and for PNMA-MS2; M. domestica (Chromosome: MonDom5: 1: 416 106 001–416 736 671), Tasmanian devil Sarcophilus harrisii (Scaffold: DEVL_7.0: GL834637:1: 1–311 016), H. sapiens (Chromosome: GRCh37: 9: 125 122 856–125 594 315), M. musculus (Chromosome: GRCh37: 2: 36 078 175–37 218 455), O. anatinus (Chromosome: OANAS: Ultra70: 222 463–282 938), G. gallus (Chromosome: WASHUC2: 17: 9 467 383–9 508 331), X. tropicalis (Scaffold: JGI4.2: GL173356:1: 228 227–292 347) and T. rubripes (Scaffold: FUGU4: scaffold_49: 144 872–154 417)

Alignments were obtained using the VISTAWeb server (http://genome.lbl.gov/vista/). PNMA-MS1 syntonic regions of several species identified above were aligned using the default setting (>70% identity and >100 bp in length) of mVISTA using the LAGAN global multiple alignment option.

3. Results

3.1. Novel candidate PNMA genes in humans and mice

We validated our approach to search for candidate PNMA genes in marsupials by performing TBLASTN analysis against human and mouse reference genomic sequences using the Gypsy12_DR Gag protein as a query. With a cut-off e-value of <1.0E−9, this screening resulted in 19 and 15 candidates in the human and mouse genomes, respectively. In humans, 15 of the 19 were known PNMA genes and the remaining 4 were novel putative PNMA genes, PNMA7/LOC649201, PNMA8/LOC649238, PNMA9/LOC100128960 and PNMA16 (Table 1, Humans). In mice, 12 of the 15 were known, with two novel putative PNMA genes, Pnma7/Gm7028 and Pnma9/Gm6858, and one pseudogene, Gm 1832215 identified (Table 1, Mouse).

The putative human PNMA7−9 genes were located near PNMA6A–D cluster on Chromosome Xq28. There is a sequence gap between PNMA6A–B and 6C–D, so additional PNMA genes may exist in this region (Supplementary Fig. S1). The high homology (47−57%) between the putative amino acid sequences of the PNMA7−9 and PNMA6A–D genes suggested that they share a common ancestor and evolved by gene duplication (see Fig. 3). The putative murine Pnma7, 8 and
| PNMA number | Gene name | Accession number | Location |
|-------------|-----------|------------------|----------|
| hsPNMA1     | PNMA1     | NM_006029.4      | chr.14:7417846-74181128 |
| hsPNMA2     | PNMA2     | NM_007257.5      | chr.8:26362196-26371483  |
| hsPNMA3     | PNMA3     | NM_013364.4      | chr.15224766-15228827   |
| hsPNMA4     | PNMA4     | NM_022151.4      | chr.14:93648541-93651249 |
| hsPNMA5     | PNMA5     | NM_052926.2      | chr.15215736-152162671  |
| hsPNMA6A    | PNMA6A    | NM_032882.4      | chr.15233801-152340107  |
| hsPNMA6B    | PNMA6B    | XM_002343859.2   | chr.15234164-152342813  |
| hsPNMA6C    | PNMA6C    | NM_01170944.1    | chr.152244152-152246070 |
| hsPNMA7     | LOC649201 | XP_001127211     | chr.152584221-152587591 |
| hsPNMA8     | LOC649238 | XM_938309.4      | chr.152662364-152663269 |
| hsPNMA9     | LOC100128960 | —         | chr.152197130-152200901 |
| hsPNMA10    | ZCCHC12   | NM_173798.2      | chr.17957787-17960931   |
| hsPNMA11    | ZCCHC18   | NM_001143978.1   | chr.103357107-103360533 |
| hsPNMA12    | PNMA12    | NM_001103149.1   | chr.14:946969748-46974820 |
| hsPNMA13    | PNMA13    | NM_020709.1      | chr.14:946994448-46999169 |
| hsPNMA14    | PNMA14    | NM_032040.3      | chr.14:94691358-46916919 |
| hsPNMA15    | PNMA15    | NM_0011100461.3  | chr.14:946931182-46931595 |
| hsPNMA16    | —         | —                | chr.19:47036933-47037357 |
| mmPNMA1     | Pnma1     | NM_027438.3      | chr.12:85487081-85489439 |
| mmPNMA2     | Pnma2     | NM_175498.4      | chr.14:67530045-67538898 |
| mmPNMA3     | Pnma3     | NM_153169.2      | chr.170310126-70313530  |
| mmPNMA4     | Pnma4     | NM_001142937.1   | chr.12:103978040-103981870 |
| mmPNMA5     | Pnma5     | NM_0011100461.3  | chr.170729732-70282442  |
| mmPNMA7     | Gm7028    | NG_005480.3      | chr.70580917-70581762   |
| mmPNMA8     | LOC100416956 | NG_017874       | chr.70642228-70644051   |
| mmPNMA9     | Gm6858    | NG_005479.2      | chr.70295221-70295900   |
| mmPNMA10    | Zcchc12   | NM_028325.3      | chr.33735899-33739153   |
| mmPNMA11    | Zcchc18   | NM_001025509.1   | chr.133527694-133531462 |
| mmPNMA12    | PNMA12    | NM_001007569.1   | chr.17545144-17547669   |
| mmPNMA13    | PNMA13    | NM_001099636.2   | chr.17530031-17532427   |
| mmPNMA14    | PNMA14    | NM_001101535.1   | chr.17579937-17581994   |
| mmPNMA15    | PNMA15    | —                | chr.175768964-17569311  |
| mmPNMA pseudo1 | Gm18322  | NC_000084.5      | chr.18:57308641-57309445 |
| Tammar      | mePNMA-MS1| —                | GeneScaffold_503:27166-31803 |
| mePNMA pseudo1 | —         | —                | Scaffold94060:2914-6134  |
| mePNMA pseudo2 | —         | —                | Scaffold1032:45408-58815 |
| mePNMA pseudo3 | —         | —                | Scaffold385831:1-1432   |
| mePNMA pseudo4 | —         | —                | Scaffold57604:970-11278 |
| mePNMA pseudo5 | —         | —                | Scaffold428007:1-986    |
| mePNMA pseudo6 | —         | —                | Scaffold1439:236-44500  |
| mePNMA pseudo7 | —         | —                | Scaffold391804:1-1797 |
| mePNMA pseudo8 | —         | —                | Scaffold46963:6594-10851 |
| mePNMA pseudo9 | —         | —                | Scaffold3242:41001-45246 |
| mePNMA pseudo10 | —        | —                | Scaffold407990:1-2479  |
| mePNMA pseudo11 | —         | —                | Scaffold799:60812-65099 |
| mePNMA pseudo12 | —         | —                | Scaffold492911:1-878    |
| mePNMA pseudo13 | —         | —                | Scaffold111753:1-2597 |
| mePNMA pseudo14 | —         | —                | Scaffold9220:2820-7152  |
| mePNMA pseudo15 | —         | —                | Scaffold9921:21555-26131 |
| mePNMA pseudo16 | —         | —                | Scaffold1103:4422-8826  |
| mePNMA pseudo17 | —         | —                | Scaffold27816:17763-21999 |
| mePNMA pseudo18 | —         | —                | Scaffold34268:3226-6573 |
| mePNMA pseudo19 | —         | —                | GeneScaffold_10085:47646-51858 |
| Opossum     | mdPNMA-MS1| mdPNMA-MS1       | —         | chr.3:260874625-260879998 |
| mdPNMA pseudo1 | —         | —                | chr.3:15383915-15388463 |
| mdPNMA pseudo2 | —         | —                | chr.3:239729898-239734143 |
| mdPNMA pseudo3 | —         | —                | chr.1:692541545-692456282 |
| mdPNMA pseudo4 | —         | —                | chr.1:451204818-451209222 |
| mdPNMA pseudo5 | —         | —                | chr.1:331217342-331221920 |
| mdPNMA pseudo6 | —         | —                | chr.1:342376567-342381307 |
| mdPNMA pseudo7 | —         | —                | chr.6503247-6507903    |
| mdPNMA pseudo8 | —         | —                | chr.2:529194963-529199310 |

Continued
9 are all located in the orthologous region on the X chromosome. However, the PNMA6 cluster is absent from the mouse genome (Supplementary Fig. S1). This region is occupied by the X-linked leucocyte-regulated complex (Xlr) gene cluster.

### 3.2. Identification of novel PNMA genes in marsupials

A comprehensive search of the tammar wallaby (Meu̇g_1.1) and opossum WGS (MonDom5) for PNMA genes was then undertaken using the same method as the human and mouse above. Twenty and 14 hits were returned for TBLASTN searches of the tammar and opossum genomes, respectively (Table 1, Tammar wallaby and Opossum). However, most of the sequences were predicted to be pseudogenes or remnants of the original retrotransposons (<100 aa).

Only ORFs predicted to encode >100 aa were considered to be marsupial PNMA candidate genes. One candidate exhibited the highest homology to the Gypsy12_DR Gag protein along with matrix (MA), N- and C-terminal parts of capsid like (N- and C-CA) and cys-cys-his-cys (CCHC) zinc finger domains and had a putative ORF consisting of 456 and 458 aa in the tammar and opossum, respectively (Fig. 1). Therefore, we named it PNMA-MS1 as a novel marsupial-specific PNMA gene. The marsupial PNMA-MS1 gene was located on a syntenic segment in the tammar (Gene scaffold_503:27168-31803) and the opossum (Chr.3: 260 874 625–260 879 998). A second PNMA candidate was identified in the opossum, PNMA-MS2, that had a putative ORF encoding 112 aa with high homology only to a central part of the capsid-like domain of the Gypsy12_DR Gag. It was located on Chromosome 1: 416 409 687–416 414 127, where an olfactory receptor (OR) gene cluster exists (see below). The presence of PNMA-MS2 in the tammar was inconclusive due to the incomplete assembly of the corresponding region of the genome.

### 3.3. Genomic structure of PNMA-MS1 in the tammar wallaby

The full-length sequence of tammar PNMA-MS1 consisting of 4290 bp was determined by 5′- and 3′-RACE. It has two exons and encodes a putative ORF encoding a 456 aa sequence (Fig. 2). The PNMA-MS1 putative ORF shared 28% similarity at the amino acid level with the Gag protein of Gypsy12_DR retrotransposon (Fig. 3A). Multiple alignment and phylogenetic tree analyses of the putative PNMA-MS1 and -MS2 amino acid sequences are shown in Fig. 3A and B. The Pol protein and LTR regions are absent from PNMA-MS1 and -MS2, suggesting that they no longer have retrotranspositional activity (Fig. 3A). The marsupial PNMA-MS1 and -MS2 protein sequences were grouped together and

| PNMA number | Gene name | Accession number | Location |
|-------------|-----------|------------------|----------|
| mdPNMA pseudo9 | — | — | chr.2:271702858-271707121 |
| mdPNMA pseudo10 | — | — | chr.5:205686598-205690882 |
| mdPNMA pseudo11 | — | — | chr.5:229359947-229364330 |
| mdPNMA pseudo12 | — | — | chr.6:103527713-103532099 |

Candidates in humans, mouse, tammar wallaby and opossum. Newly identified PNMA family genes in this study are coloured in grey. ‘Pseudo’ denotes putative ORFs from sequences detected by TBLASTN, which encode <100 aa.
Figure 3. Multiple sequence alignment and phylogenetic tree of the PNMA family. (A) Multiple sequence alignment of the amino acid sequence of the Gag-like regions of marsupial PNMA-MS1, the human and mouse PNMA genes and Gypsy12_I_DR Gag. An evolutionarily conserved Gag-derived CX2CX4HX4C zinc finger motif is indicated by yellow shading. Residues conserved in all sequences are shaded black and highly conserved ones in grey. (B) A phylogenetic tree of PNMA family genes was constructed by the neighbour-joining method using the multiple alignment shown in Fig. 2. Bootstrap support (%) is shown for branches. hs: human; mm: mouse; md: opossum; me: wallaby.
more closely related to the zebrafish Gypsy12_DR Gag than to the mouse and human proteins (Fig. 3B).

3.4. Comparative genomic analysis of PNMA-MS1 and -MS2

To elucidate whether PNMA-MS1 is a marsupial-specific PNMA gene, comparative genomic analysis was performed using the VISTA tool with several vertebrate genomic sequences. PNMA-MS1 was located in the intron 8 of the laminin alpha 3 (LAMA3) gene that is highly conserved in vertebrates. No PNMA-MS1 orthologue was found in the syntenic region of any eutherian species, platypus (monotreme mammals), chicken (birds), frog (amphibian) and fugu (fish), demonstrating that PNMA-MS1 is marsupial specific (Fig. 4A). It indicates that PNMA-MS1 retrotransposition occurred only in the marsupial lineage after their divergence from eutherians (Fig. 5).

PNMA-MS2 was located between an OR1Q1 gene and an OR1J2-like pseudogene (ENSMODG0000019710) that lies 12-kb upstream of the former, in an OR gene cluster located between prostaglandin-endoperoxide synthase 1 (PTGS1) and phosducin-like (PDCL) genes on opossum Chromosome 1. At present, we cannot confirm the presence or absence of PNMA-MS2 in the tammar because the corresponding regions encompassing the PTGS1 and PDCL genes are yet to be completely assembled. Recently, another Australian marsupial genome of the Tasmanian devil has been sequenced, and the region syntenic to that between opossum OR1N2 and PDCL became available.

Figure 3. Continued
Figure 4. Comparative genomic analysis of the PNMA-MS1 and -MS2 regions in vertebrates. (A) PNMA-MS1. mLAGAN alignment of the tammar wallaby, opossum, human, mouse, platypus, chicken, frog and fugu LAMA3 exons 8–10 region produced by mVISTA using the tammar sequence as the basis for comparison. Default parameters for mVISTA were used (conservation level, 70%, 100 bp window). Conserved regions appear as peaks highlighted in pink (>70% identity). Where these regions coincide with ORF sequences of PNMA-MS1 or LAMA3, the peaks are shaded in purple. Where these regions coincide with the UTR region of PNMA-MS1, the peaks are shaded in light blue. The
It contains some gaps but none between opossum OR1Q1 and ENSMODG00000019710 (OR1J2-like pseudogene). A search of this region clearly demonstrated that the PNMA-MS2 orthologue is absent from the Tasmanian devil genome.

The syntenic OR cluster lies between PTGS1 and PDCL in the human Chromosome 9 and mouse Chromosome 2, respectively, but neither the number nor the order of OR genes and pseudogenes are conserved. As the human and mouse genome sequences in this region are complete and contain no gaps, the absence of the PNMA-MS2 orthologue was confirmed (Fig. 4B). For the platypus, the PTGS1 and PDCL genes are located next to each other, with no OR gene cluster and no PNMA-MS2 orthologue, like in the chicken and fugu (Fig. 4B).

These results suggest that the integration of selected OR genes occurred between the PTGS1 and PDCL genes in a common therian ancestor, and that the opossum-specific insertion of PNMA-MS2 occurred after the divergence of eutherians and marsupials and the geographic separation of Australian and South American marsupials (Fig. 5). However, the possibility that the PNMA-MS2 orthologue exists in some of Australian marsupials was not excluded. It is possible that PNMA-MS2 was deleted from Australian marsupial species after integration in a common marsupial ancestor (Fig. 5).

3.5. Expressions of PNMA-MS1 in the tammar wallaby and PNMA-MS2 in the opossum

PNMA-MS1 expression was investigated in several tissues in four different stages of the tammar wallaby, including foetal and pouch young stages. Human LAMA3 is expressed ubiquitously (EST profile Hs.436367). To exclude the possibility that heterogenous nuclear RNA (hnRNA) was detected between exons 8 and 9 of the LAMA3 gene rather than PNMA-MS1, we amplified PNMA-MS1 using PCR primers designed within exons 1 and 2, respectively. Thus, the PCR product was shorter than its genomic sequence corresponding to hnRNAs of LAMA3 and PNMA-MS1. LAMA3 expression was analysed using primers designed to exons 79 and 81, near the 3'-UTR, due to the poor genome sequence quality in introns 84–90 of tammar LAMA3. Tammar LAMA3 expression was almost ubiquitous, with the exception of several pouch young tissues.

From Day 23 to 26 pregnancy, PNMA-MS1 expression was detected in the foetal head and body, but there was no expression in the YSP (Fig. 6A). In pouch young aged Day 60–70, PNMA-MS1 was detected only in the brain, kidney and ovary, but not in the liver, lung or testis (Fig. 6B). In Day 152 and 162 pouch young, changes in PNMA-MS1 expression were minimal, with expression detected in the kidney, liver, pancreas, heart, spleen and stomach, but not in the lung, bladder, adrenal, cerebrum or cerebellum (Fig. 6C). In the adult female, the brain (thalamus, hypothalamus and pituitary), ovary (ovary with active corpus luteum, corpus luteum alone, ovary with enlarged developing follicle and ovary with primary or secondary follicles), endometrium (gravid endometrium and non-gravid endometrium) and mammary gland (sucked gland and non-sucked gland) were examined. PNMA-MS1 expression was detected in the ovary (all four stages), mammary gland (sucked and non-sucked) and thalamus, but not in the hypothalamus or pituitary (Fig. 6D). There was no expression in either the gravid, or non-gravid, endometrium.

PNMA-MS2 expression was analysed in five tissues, the brain, liver, spleen, pituitary and ovary, using three primer sets designed in the putative coding frame. However, there was no expression in these tissues, suggesting that PNMA-MS2 is not active in the opossum (data not shown). Although we cannot exclude the possibility that it may be expressed in a stage or tissue-specific manner, we conclude that PNMA-MS2 is a pseudogene.

4. Discussion

4.1. PNMA-MS1 is a marsupial-specific PNMA gene

In this study, we have identified PNMA-MS1 as a novel Ty3/Gypsy LTR retrotransposon-derived gene. Comparative genomic analysis showed that PNMA-MS1 was present only in the marsupial lineage and was absent in the eutherian and monotreme mammals and in the non-mammalian vertebrates. PEG10/SIRH1 and SIRH12 are the only Ty3/Gypsy LTR retrotransposon-derived genes (derived from sushi-ichi-related retrotransposons) reported in marsupials so far.\(^\text{11,21}\) Therefore, the PNMA-MS1, which is also a Ty3/gypsy LTR retrotransposon-derived gene, is the first and only member of the PNMA gene family in the marsupials.

The sushi-ichi-related retrotransposon that gave rise to the SIRH family was probably active around the...
time of the divergence between marsupials and eutherians, because \( \text{PEG10}/\text{SIRH1} \) is conserved between the eutherians and marsupials,\(^{21}\) while \( \text{SIRH3–11} \) seems to be eutherian-specific and \( \text{SIRH12} \) evolved from a marsupial-specific retrotransposition event.\(^{11}\) We did not detect any orthologues of \( \text{PNMA-MS1} \) or \( \text{MS2} \) in the eutherian genome, nor any orthologues of eutherian \( \text{PNMA1–16} \) genes in the marsupial genomes (data not shown). Due to many sequence gaps in the tammar wallaby, Tasmanian devil and opossum genomes, we cannot exclude the possibility that some marsupial orthologues of eutherian \( \text{PNMA} \) genes exist in such gap regions. Thus, it is possible that some retrotransposition events of Gypsy12_DR-related retrotransposon occurred in the common ancestor of the marsupials and eutherians. However, the higher similarity of \( \text{PNMA-MS1} \) and \( \text{MS2} \) to Gypsy12_DR Gag than any other eutherian \( \text{PNMAs} \) suggests that their insertions in the marsupial genome were recent events. Taken together, these results suggest that the retrotransposition of Gypsy12_DR-related retrotransposon occurred after the divergence of the marsupials and eutherians. The \( \text{PNMA} \) genes then evolved independently in these two lineages (Fig. 5).

In our analysis, only \( \text{PNMA-MS1} \) was detected in the tammar and opossum in contrast to 19 and 14 \( \text{PNMA} \) genes in humans and mice, respectively. We also observed the same trend in the \( \text{SIRH} \) genes: 11 genes in both humans and mice, while there are only 2 genes (\( \text{PEG10} \) and \( \text{SIRH12} \)) in the tammar genome.\(^{11}\) This implies that the eutherian genome has a greater ability of exaptation as more Ty3/Gypsy types of LTR retrotransposons were incorporated into the genomes as endogenous genes than in the marsupial genomes.

![Figure 5. Evolutionary pathway of the PNMA family in mammals.](image)

### 4.2. The possible role of PNMA-MS1 genes in marsupial development

In rare cases, some retrotransposons have been incorporated as novel acquired genes into the host genomes and have contributed to the innovation of some eutherian-specific characteristics. Two such advantageous genes, \( \text{PEG10}/\text{SIRH1} \) and \( \text{PEG11}/\text{SIRH2} \), play essential roles in the placental development in mice.\(^{7,8}\) The role of \( \text{PNMA-MS1} \) in marsupial development and growth is less clear. \( \text{PNMA-MS1} \) expression was detected in the tammar brain, consistent with the expression of eutherian \( \text{PNMA} \) genes in brain. Interestingly, \( \text{PNMA-MS1} \) expression was confirmed only in the thalamus, but not in the hypothalamus, and pituitary in the adult brain. The thalamus has multiple functions including relaying sensation, spatial sense and motor signals to the cerebral cortex,\(^{22}\) so it is possible that \( \text{PNMA-MS1} \) is involved in the transmission of marsupial-specific sensations and signals. \( \text{PNMA-MS1} \) expression in the Day 60–70 pouch young ovary and adult female ovary suggests the gene may have a role in ovarian function. These issues will be addressed in a future study.

\( \text{PNMA-MS1} \) protein has a conserved CCHC zinc finger domain. In retroviruses, this domain forms a part of the nucleocapsid protein that functions in virus genome packaging and the early infection process.\(^{23}\) Proteins containing the CCHC zinc finger domain are commonly known to interact with single-stranded DNAs (ssDNAs) and RNAs.\(^{24}\) The \textit{Drosophila} Nanos protein is required for \textit{hunchback} mRNA translational regulation in the early embryo to the establishment of the anterior–posterior body axis.\(^{25}\) The mammalian cellular nucleic acid-binding protein (CNBP) containing seven CCHC domains is involved in neural crest development, affecting forebrain and craniofacial development.\(^{26}\) CNBP has a single-stranded nucleic acid-binding ability and is also implicated in both transcriptional and translational regulations.\(^{27,28}\) Therefore, \( \text{PNMA-MS1} \) may also be involved in a specific-transcriptional or translational regulation by binding ssDNAs or RNAs.
5. Conclusions

We have identified one novel Ty3/Gypsy LTR retrotransposon-derived gene, PNMA-MS1 in marsupials as the first marsupial-specific PNMA gene reported. The high PNMA-MS1 expression levels in the thalamus, ovary, and mammary gland provide intriguing questions as to its functions in marsupial development and growth as well as its role in marsupial evolution. Our data suggest that, in most of the cases, Ty3/Gypsy LTR retrotransposons have been independently incorporated into the marsupial and eutherian lineages. Consequently, the marsupials and eutherians have completely different sets of PNMA and SIRH genes, with the exception of PEG10. Thus, it is highly likely that these genes have evolved lineage-specific functions in the reproduction and development and contributed in establishing marsupial- or eutherian-specific traits, leading to the diversification of these two viviparous mammalian groups.

Acknowledgements: We thank Scott Brownlees, Bonnie Dopheide, Hongshi Yu, Jessica Stringer, Stephen Frankenberg and the Wallaby research group for the assistance with the animals and tissue collections, and Soichi Ogishima for his advice on phylogenetic analysis.

Supplementary Data: Supplementary Data are available at www.dnaresearch.oxfordjournals.org.

Accession Numbers

The National Center for Biotechnology Information GenBank (http://www.ncbi.nlm.nih.gov/Genbank) sequence accession number for tammar PNMA-MS1 is AB646689.

Funding

This work was supported by grants from the Australian Research Council Centre of Excellence in Kangaroo Genomics to M.B.R., Japan Society for the Promotion of Science (JSPS), Bilateral Program on Joint Research Project to F.I. and T.K-I. and Funding Program for Next Generation World-Leading Researchers (NEXT Program) to T.K-I., the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan, Grants-in-Aid for Scientific Research (S) and the
Global Center for Excellence (GCOE) Program for International Research Center for Molecular Science in Tooth and Bone Diseases and Medical Research Institute, Tokyo Medical and Dental University, Joint Usage/Research Program to FI.

References

1. Lander, E., Linton, L., Birren, B., et al. 2001, Initial sequencing and analysis of the human genome, Nature, 409, 860–921.
2. Mouse Genome Sequencing Consortium Waterston, R.H., Lindblad-Toh, K., Birney, E., et al. 2002, Initial sequencing and comparative analysis of the mouse genome, Nature, 420, 520–62.
3. Warren, W., Hillier, L., Marshall Graves, J., et al. 2008, Genome analysis of the platypus reveals unique signatures of evolution, Nature, 453, 175–83.
4. Mikkelsen, T., Wakefield, M., Aken, B., et al. 2007, Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences, Nature, 447, 167–77.
5. Renfree, M., Papenfuss, A., Deakin, J., et al. 2011, Genome sequence of an Australian kangaroo, *Macropus eugenii*, provides insight into the evolution of mammalian reproduction and development, Genome Biol., 12, R81.
6. Miller, K., Lynch, C., Martin, J., Herniou, E. and Tristem, M. 1999, Identification of multiple GypsyLTR-retrotransposon lineages in vertebrate genomes, J. Mol. Evol., 49, 358–66.
7. Ono, R., Nakamura, K., Inoue, K., et al. 2006, Deletion of Peg10, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality, Nat. Genet., 38, 101–6.
8. Sekita, Y., Wagatsuma, H., Nakamura, K., et al. 2008, Role of retrotransposon-derived imprinted gene, Rtl1, in the feto-maternal interface of mouse placenta, Nat. Genet., 40, 243–8.
9. Brandt, J., Veith, A. and Volff, J. 2005, A family of neofunctionalized Ty3/gypsy retrotransposon genes in mammalian genomes, Cytogenet. Genome Res., 110, 307–17.
10. Youngson, N., Kociaklowksi, S., Pearl, N. and Ferguson-Smith, A. 2005, A small family of sushi-class retrotransposon-derived genes in mammals and their relation to genomic imprinting, J. Mol. Evol., 61, 481–90.
11. Ono, R., Kuroki, Y., Naruse, M., et al. 2011, Identification of tammar wallaby SRH12, derived from a marsupial-specific retrotransposition event, DNA Res., 18, 211–9.
12. Kaneko-Ishino, T. and Ishino, F. 2012, The role of genes domesticated fromLTRretrotransposons and retroviruses in mammals, Front Microbiol., 3, 262.
13. Campillos, M., Doerks, T., Shah, P. and Bork, P. 2006, Computational characterization of multiple Gag-like human proteins, Trends Genet., 22, 585–9.
14. Takaji, M., Komatsu, Y., Watakabe, A., Hashikawa, T. and Yamamori, T. 2009, Paraneoplastic antigen-like 5 gene (PNM5) is preferentially expressed in the association areas in a primate specific manner, Cereb. Cortex, 19, 2865–79.
15. Rosenfeld, M., Eichen, J., Wade, D., Posner, J. and Dalmau, J. 2001, Molecular and clinical diversity in paraneoplastic immunity to Ma proteins, Ann. Neurol., 50, 339–48.
16. Schüller, M., Jenne, D. and Voltz, R. 2005, The human PNMA family: novel neuronal proteins implicated in paraneoplastic neurological disease, J. Neuroimmunol., 169, 172–6.
17. Poole, WE., Simms, N.G., Wood, J.T. and Lubulwa, M. 1991, *Tables for Age Determination of the Kangaroo Island Wallaby (Tammar),* Macropus eugenii, *from Body Measurements*. Technical Note. Division of Wildlife and Ecology, CSIRO: Canberra, Australia.
18. NHMRC 2004, *Australian Code of Practice for the Care and Use of Animals for Scientific Purposes.* 7th edition. Australian Government: Canberra.
19. Godfrey, P., Malnic, B. and Buck, L. 2004, The mouse olfactory receptor gene family, Proc. Natl Acad. Sci. USA, 101, 2156–61.
20. Malnic, B., Godfrey, P. and Buck, L. 2004, The human olfactory receptor gene family, Proc. Natl Acad. Sci. USA, 101, 2584–9.
21. Suzuki, S., Ono, R., Narita, T., et al. 2007, Retrotransposon silencing by DNA methylation can drive mammalian genomic imprinting, PLoS Genet., 3, e55.
22. Ashwell, K. 2010, The neurobiology of Australian marsupials: brain evolution in the other mammalian radiation, Cambridge University Press, pp. 95–6.
23. Narayanan, N., Gorelick, R. and DeStefano, J. 2006, Structure/function mapping of amino acids in the N-terminal zinc finger of the human immunodeficiency virus type 1 nucleocapsid protein: residues responsible for nucleic acid helix destabilizing activity, Biochemistry, 45, 12617–28.
24. Matsu, T., Kodera, Y., Endoh, H., et al. 2007, RNA recognition mechanism of the minimal active domain of the human immunodeficiency virus type-2 nucleocapsid protein, J. Biochem., 141, 269–77.
25. Curtis, D., Treiber, D., Tao, F., Zamore, P., Williamson, J. and Lehmann, R. 1997, A CCHC metal-binding domain in Nanos is essential for translational regulation, EMBO J., 16, 834–43.
26. Chen, W., Liang, Y., Deng, W., et al. 2003, The zinc-finger protein CNBP is required for forebrain formation in the mouse, Development, 130, 1367–79.
27. Rajavashisth, T., Taylor, A., Andalibi, A., Svenson, K. and Lusis, A. 1989, Identification of a zinc finger protein that binds to the sterol regulatory element, Science, 245, 640–3.
28. Schlatter, S. and Fussenegger, M. 2003, Novel CNBP- and La-based translation control systems for mammalian cells, Biotechnol. Bioeng., 81, 1–12.
