Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Mamm Genome. Author manuscript; available in PMC 2016 April 17.

Published in final edited form as:

Mamm Genome. 2012 June ; 23(5-6): 336–345. doi:10.1007/s00335-011-9383-x.

Bridging sequence diversity and tissue-specific expression by DNA methylation in genes of the mouse prolactin superfamily

Koji Hayakawa, **Momo O. Nakanishi**, **Jun Ohgane**, **Satoshi Tanaka**, **Mitsuko Hirosawa**, **Michael J. Soares**, **Shintaro Yagi**, and **Kunio Shiota**

Laboratory of Cellular Biochemistry, Department of Animal Resource, Sciences/Veterinary Medical Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan

Abstract

Much of the DNA in genomes is organized within gene families and hierarchies of gene superfamilies. DNA methylation is the main epigenetic event involved in gene silencing and genome stability. In the present study, we analyzed the DNA methylation status of the *prolactin* (*PRL*) superfamily to obtain insight into its tissue-specific expression and the evolution of its sequence diversity. The *PRL* superfamily in mice consists of two dozen members, which are expressed in a tissue-specific manner. The genes in this family have CpG-less sequences, and they are located within a 1-Mb region as a gene cluster on chromosome 13. We tentatively grouped the family into several gene clusters, depending on location and gene orientation. We found that all the members had tissue-dependent differentially methylated regions (T-DMRs) around the transcription start site. The T-DMRs are hypermethylated in nonexpressing tissues and hypomethylated in expressing cells, supporting the idea that the expression of the *PRL* superfamily genes is subject to epigenetic regulation. Interestingly, the DNA methylation patterns of T-DMRs are shared within a cluster, while the patterns are different among the clusters. Finally, we reconstituted the nucleotide sequences of T-DMRs by converting TpG to CpG based on the consideration of a possible conversion of 5-methylcytosine to thymine by spontaneous deamination during the evolutionary process. On the phylogenic tree, the reconstituted sequences were well matched with the DNA methylation pattern of T-DMR and orientation. Our study suggests that DNA methylation is involved in tissue-specific expression and sequence diversity during evolution.

Introduction

Gene duplication is one of the mechanisms of creating new genetic material, thereby contributing to evolution (Louis 2007). The duplicated genes, once copied in the genome, have a number of possible fates, including gaining a novel function or a new regulatory

ashiota@mail.ecc.u-tokyo.ac.jp.

Present Address: J. Ohgane Laboratory of Genomic Function Engineering, Department of Life Sciences, School of Agriculture, Meiji University, 1-1-1 Higasimita, Tama-ku, Kawasaki 214-8571, Japan

Present Address: M. J. Soares Institute for Reproductive Health and Regenerative Medicine, Department Pathology and Laboratory Medicine, University of Kansas Medical Center, Kansas City, KS 66160, USA

Electronic supplementary material The online version of this article (doi:10.1007/s00335-011-9383-x) contains supplementary material, which is available to authorized users.

Conflicts of interest The authors have no conflicts of interest to declare.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

mechanism that allows changes in the level and/or pattern of expression via sequence divergence and chromosomal localization. During evolution, epigenetic changes may also participate in modifying the mutation rate, through C-to-T transitions due to deamination of methylcytosines and through modifying rates of chromosomal rearrangement (Janion 1982; Carbone et al. 2009).

DNA methylation constitutes an important level of epigenetic control and has been implicated in the control of tissue-specific gene expression in mammals (Imamura et al. 2001; Li 2002; Shiota 2004). We previously showed that the rat placental lactogen gene *PL-I* has a tissue-dependent, differentially methylated region (T-DMR) and that placenta-specific gene expression is regulated by DNA methylation (Cho et al. 2001). Genome-wide DNA methylation analysis of normal tissues and cells indicated the widespread presence of T-DMRs around the transcription start sites of various types of genes carrying not only CpG islands but also CpG-poor promoters, including transcription factors and their targets in normal tissues and cells (Shiota et al. 2002; Rakyan et al. 2008; Yagi et al. 2008).

The *prolactin* (*PRL*) superfamily consists of 24 closely related genes in mice and is located within a 1 Mb region on chromosome 13 (Wiemers et al. 2003; Mallon et al. 2004; Simmons et al. 2008). The *PRL* superfamily genes are categorized into eight subfamilies in the mouse genome database (Soares et al. 2007a). The *PRL* members have diverse functions and cell- or tissue-specific gene expression patterns. *Prl* is expressed primarily in the pituitary, and the remaining genes encoding *PLs* and *PRL-like proteins* (*PLPs*) are expressed mainly in the uterine decidua and placenta (Soares 2004; Soares et al. 2007b). Expression of some *PRL* superfamily members is observed in normal (Linzer and Nathans 1984; Fassett and Nilsen-Hamilton 2001; Choong et al. 2003) or transformed fibroblasts/mesenchymal cells (Toft et al. 2001), and can be activated under pathological conditions in hematopoietic cells (Bhattacharyya et al. 2002; Ho-Chen et al. 2007).

In mammalian genomes, CpGs are present at one-fifth of their expected frequency due to the spontaneous conversion of 5-methylcytosine into thymine by deamination (Russell et al. 1976; Coulondre et al. 1978). The *PRL* superfamily genes have a relatively low frequency of CpGs, and the number of CpGs varies among the superfamily genes around the transcription start sites. We hypothesized that DNA methylation may be a possible mechanism to cause mutation, producing variations in the number of CpGs. Indeed, the number of CpGs is low in rat *PL-I*, which has a T-DMR (Cho et al. 2001). It is not known, however, if there are T-DMRs in the superfamily members.

In this study we explored the connection between how diverse superfamily gene members are formed and how the tissue-specific expression patterns are influenced by DNA methylation status. We used the T-DMR profiling with restriction tag-mediated amplification (D-REAM) and the bisulfite method to investigate the methylation status of the mouse *PRL* superfamily genes.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Materials and methods

Animals and tissue preparation

Adult mice (C57BL/6N) were purchased from Charles River Japan (Yokohama, Japan) and were maintained on a 12 h light/12 h dark schedule with free access to food and water. When a vaginal plug was observed at noon of a given day, tissue samples were designated as embryonic day 0.5 (d 0.5). The placenta at d 14.5 was recovered by separation from the decidua using fine forceps. The liver, cerebrum, and pituitary were dissected from 13-week-old male mice. Collected tissues were stored at −80°C until used for DNA extraction.

All experiments using mice were carried out according to the institutional guidelines for the care and use of laboratory animals (Graduate School of Agriculture and Life Sciences, The University of Tokyo).

DNA methylation analysis by the bisulfite method

Genomic DNA was extracted from tissues as described previously (Yagi et al. 2008). Briefly, each sample was treated in lysis buffer (10 mM Tris-HCl, pH 8.0, 5 mM EDTA, 200 mM NaCl, 0.2% SDS, and 200 μg/ml proteinase K) following phenol/chloroform/isoamyl alcohol (50:49:1) extraction and dissolved in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0).

Genomic DNA, digested with *Eco*RI or *Pst*I, was denatured by adding 0.3 M NaOH and incubating for 15 min at 37°C. After incubation, sodium metabisulfite (pH 5.0), and hydroquinone were added to final concentrations of 2.0 M and 0.5 mM, respectively, and the mixture was further incubated at 55°C for 16 h in the dark. The bisulfite reaction was terminated by purification using a Qiagen gel extraction kit (Qiagen, Valencia, CA, USA) followed by incubation in NaOH (0.3 M, final) at 37°C for 15 min. The modified DNA was ethanol-precipitated after neutralization with NH_4OAc (pH 7.0) at a final concentration of 3 M. The ethanol-precipitated DNA was dissolved in TE buffer and then amplified by PCR using primers designed for sequencing and combined bisulfite restriction analysis (COBRA) (Supplementary Table 1).

For COBRA, PCR fragments were digested with HpyCH4IV (New England Biolabs, Ipswich, MA, USA) to evaluate the methylation status of significantly different HpyCH4IV sites. Because only unmethylated cytosine residues are changed to thymines by the sodium bisulfite reaction followed by PCR, PCR fragments from nonmethylated genomic DNA are resistant to HpyCH4IV and those from methylated DNA are digested by the enzyme. For sequencing, the PCR fragments were cloned into pGEM T-Easy vector (Promega, Madison, WI, USA), and ten clones were sequenced for each sample.

D-REAM analysis

We applied D-REAM analysis, as described previously (Yagi et al. 2008), to examine tissue-dependent and differentially methylated HpyCH4IV sites around the transcription start site (−6 to +2.5 kb) of *PRL* superfamily genes. Genomic DNA (50 ng) was digested by HpyCH4IV and analyzed using D-REAM. Different methylation levels were analyzed using

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

ligation-mediated PCR and hybridization with mouse promoter arrays around the digested HpyCH4IV sites. In this study we performed D-REAM analysis for the placenta, liver, and cerebrum. For the liver and cerebrum, we used signal D-REAM data from our previous publication (Yagi et al. 2008). In D-REAM analysis, differences of methylation status, namely, T-DMR, were considered significant if $P < 10^{-3}$.

Phylogenetic analysis

cDNA and genomic [±1 kb from transcription start sites (TSS)] sequences of the 24 paralogous *PRL* family members were aligned into the CLUSTAL W and TREEVIEW program to generate a phylogenetic tree (Thompson et al. 1994).

Results

Genomic characteristics of the mouse *PRL* superfamily member genes on chromosome 13

In this study we divided the mouse *PRL* superfamily members, consisting of 24 genes that are expressed in a tissue-specific manner, into four clusters according to their genomic location and orientation of transcription (Table 1). Each cluster consisted of several tandemly repeated family genes that had been classified according to homogeneity of coding sequences, function (classical/nonclassical), or gene exon/intron structure (alias symbols) (Soares 2004; Soares et al. 2007a). Moreover, the genes in a group are transcribed in the same direction, except for cluster D, which consists of family members that have inverted orientation. For example, cluster A has nine genes of four subfamily members, transcribed from centromere to telomere, and cluster B has ten genes of three subfamily members, transcribed from telomere to centromere. In both clusters there are tandemly repeated predominant family members: four *Prl3* members in cluster A, and four *Prl8* and five *Prl7* members in cluster B. In contrast, cluster C has single copies of three different subfamilies.

We then examined sequence characteristics within 2 kb genomic regions (from −1 kb to +1 kb of the TSS) of genes in these four clusters. The number of CpGs varies among the genes in any cluster (Table 1 and Fig. 1a). For example, within the 2 kb region that includes *Prl*, there are only 4 CpGs, while there are 12 CpGs within the 2 kb region surrounding *Prl3d1*. Even in the CpG-abundant genes such as *Prl3d1* and *Prl6a1*, the ratio of observed to expected CpG dinucleotide frequencies (CpGobs/CpGexp), indicating CpG density, is quite low, suggesting that the *PRL* superfamily members had been under evolutionary pressure to reduce CpG frequency. It is noteworthy that intergenic regions in these clusters exhibited long interspersed elements (LINE)-rich and short interspersed repeated DNA elements (SINE)-poor genomic characteristics (Supplementary Fig. 1). The unidirectional tandem repetitive gene clusters at chromosome 13, with reduced and varied frequencies of CpGs, are unique characteristics of *PRL* superfamily genes.

Tissue-specific DNA methylation status of *Prl* and *Prl3d1*

Prl and *Prl3d1* in cluster A differ from each other in their tissue-specific expression patterns and their CpG densities. *Prl* is expressed primarily in the pituitary with lower levels of expression in differentiated uterine stromal cells and potentially other tissues, whereas *Prl3d1* is expressed exclusively in trophoblast giant cells of the placenta (Soares 2004;

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Soares et al. 2007b). In the analysis of *Prl3d1*, bisulfite sequencing revealed that eight CpGs downstream (+98 to +799) and three CpGs upstream (−238 to −64) were hypomethylated in the placenta and hypermethylated in the liver, cerebrum, and pituitary. In contrast to *Prl3d1*, the sequences surrounding *Prl* contain only a few CpGs; however, bisulfite sequencing indicated that the three CpGs at the upstream region (−643 to −98) of *Prl* were hypomethylated specifically in the pituitary (Fig. 1b). In contrast, these CpGs were hypermethylated in the placenta, liver, and cerebrum. Thus, the T-DMRs at *Prl* and *Prl3d1* are hypermethylated in nonexpressing tissues and hypomethylated in expressing tissues.

D-REAM and bisulfite sequencing identified the T-DMRs at all *PRL* superfamily member genes

We next explored T-DMRs in other members of the *PRL* superfamily genes. We applied D-REAM to detect T-DMRs in the family genes and found T-DMRs hypomethylated in the placenta compared with the liver and cerebrum. The T-DMRs were located in sequences from −6 kb to +2.5 kb of the TSS in five genes of cluster A (*Prl3d2, Prl3b1, Prl3a1, Prl6a1,* and *Prl8a2*) (Fig. 2a, Supplementary Fig. 2A). Since there were no HpyCH4IV sites in *Prl3d3* and *Prl3c1*, we further analyzed the DNA methylation of cluster A members by combined bisulfite restriction analysis (COBRA) and bisulfite sequencing (Fig. 2a, Supplementary Fig. 2A). T-DMRs found in members of cluster C (*Prl4a1*, *Prl5a1*, and *Prl2a1*) exhibited tissue-dependent differential methylation patterns similar to those found in cluster A members (Fig. 2a, b, Supplementary Fig. 2C). These data indicated that the T-DMRs, most of which were expanded on both sides of their TSS, were hypomethylated in the placenta and hypermethylated in the liver and cerebrum in clusters A and C (Fig. 2b).

The T-DMRs in the members of clusters B and D exhibited different methylation patterns from those in clusters A and C; they were partially hypomethylated in the liver and hypomethylated in the placenta (Fig. 2b, Supplementary Fig. 2B, D). The *PRL* gene clusters, classified by their genomic location and orientations, could be regarded as epigenomic clusters exhibiting similar DNA methylation profiles.

Sequence properties of *PRL* superfamily genes and their evolution

In cluster of *PRL* genes, epigenetic status and gene expression patterns were closely associated with their localization on genome. Genes in clusters A and C are expressed in placenta and/or decidua, except *Prl* (Table 1). However, *Prl7a1* and *Prl2c* in clusters B and D, respectively, are expressed in placenta and other tissues but not decidua. These results suggested a common regulatory mechanism in a cluster and prompted us to explore genomic features associated with these clusters. When the clusters were grouped by orientation of transcription, the number of TpGs and CpAs of genes transcribed from centromere to telomere were significantly higher compared with those transcribed in the opposite direction (Fig. 3b, c). On the other hand, the number of CpGs located within −1 kb from the TSS of genes transcribed from centromere to telomere was significantly lower compared that of with genes transcribed in the opposite direction (Fig. 3a). Thus, the genomic sequences of genes on the positive strand were CpG-poor and TpG/CpA-rich compared with those of genes on the negative strand. The DNA methylation patterns of T-DMRs were similar among the genes grouped by orientation, with the exception of *Prl2c5*.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

A phylogenetic tree based on the genomic sequences spanning the 2-kb regions around the TSS where T-DMRs are located does not exhibit any gene clusters (Fig. 4a). In contrast, a phylogenetic tree based on cDNA sequences can classify *PRL* genes into gene families as proposed (Supplementary Fig. 3). A lower frequency of CpGs in *PRL* family genes suggested that extensive C-T conversion may have occurred in them; therefore, we hypothesized that a comparison of the putative ancestral (reconstituted) genomic sequences could be carried out using the sequences obtained by converting TpG into CpG in these regions. When we converted TpG–CpG in the aligned genomic sequences, the genes were categorized into the clusters of gene families based on their coding sequences. However, the genes in the *Prl2* and *Prl8* subfamilies were separated into different groups: one group contained a cluster of *Prl8a2, Prl2c5,* and *Prl6a1*, one group contained *Prl2c3, Prl2b1,* and *Prl7* family genes, and another group contained *Prl4a1, Prl2a1,* and *Prl5a1* (Fig. 4b). It is worth noting that cluster B contained *Prl7* members and two members of *Prl2*. The orientation and DNA methylation profiles of these genes were similar to each other with the exception of *Prl2c5*, which exhibited different methylation patterns from other members on the positive strand.

Discussion

The *PRL* superfamily genes, present in large clusters on chromosome 13, can be classified into four clusters according to their position and orientation. Our results showed that similar DNA methylation patterns were exhibited at the identified T-DMRs within a classified gene cluster; thus, the unit of genomic cluster coincided with that of the epigenomic cluster. The genes in a cluster showed similar expression patterns in terms of tissue specificity, which was negatively correlated with DNA methylation levels. These T-DMRs were hypermethylated in nonexpressing tissues and hypomethylated in expressing cells. In our previous study, we showed that hypomethylation around TSS of genes is the permissive state of expression and hypermethylation around TSS is the silencing state (Yagi et al. 2008; Sato et al. 2010), and methylation of promoters forces repression of gene expression (Hattori et al. 2004; Tomikawa et al. 2006). Taken together, DNA methylation should ensure tissue-dependent stringent regulation of *PRL* genes.

Reconstituted genomic sequences of 2 kb around the TSS, which were deduced by conversion of TpGs into CpGs, exhibited similarities of sequences within the family members in clusters A and B, suggesting that tandem gene clusters may have been generated by gene duplication of a member in the repeats. This hypothesis coincides with previous analyses suggesting that the cluster of *PRL* genes in rodents may have resulted from gene duplications, not segmental duplications, based on their coding sequences and exon-intron structures (Soares 2004). Our data indicated that DNA methylation patterns and gene expression patterns might be associated with most of the duplication process in the cluster. Another gene duplication pattern observed in the genes in cluster C is intriguing. Alignment of the reconstituted genomic sequences suggests that these genes may have been derived from a common sequence and then mutated into different family genes to gain a function while retaining their epigenomic information. Based on these results, we propose that the classification by reconstitution is a better representation of the evolution of the *PRL* cluster.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Histone modification represents another form of epigenetic regulation. The human growth hormone (GH) gene cluster shares an ancestral gene with the *Prl* superfamily (Ho et al. 2004; Kimura et al. 2007). Transposable elements such as Alu and P elements are targets for histone modification and contribute to the regulation of the human GH gene cluster. Mouse *PRL* clusters exhibit rather distinct genomic features, containing CpG-poor, LINE-rich, and SINE-poor genomic features, which are common characteristics of genomic regions associated with hypermethylated T-DMRs in mouse ES cells (Muramoto et al. 2010). C-T conversion at the methylated T-DMRs in duplicated genomic regions would provide clues to the mechanisms involved in spatial and temporal expression of *PRL* family genes in mouse placenta.

The reconstituted genomic sequences of *Prl2c* genes located in cluster D show similarities with family members in other clusters, and one of them, *Prl2c5*, exhibited a different methylation profile from those genes with similar reconstituted sequences. The members of the *Prl2* sub-families, located in clusters B and C, exhibited similarities to other family members in the cluster B and C and also exhibited similar methylation profiles. These data suggest that the epigenomic domain defined by epigenetic clusters could dictate the DNA methylation profile of the integrated genes, and the localization change could provide an opportunity for altering epigenetic regulation, resulting in the alteration of gene expression patterns. Such epigenetic domains are identified by epigenomic analysis of mouse ES cells, including biased enrichment of SINEs (Muramoto et al. 2010).

The numbers of CpGs and TpG/CpAs were different, not only within the clusters but also among the clusters. Genes in a cluster exhibited similar expression profiles and DNA methylation patterns; however, some genes had different expression patterns. DNA methylation has been proposed to be involved in genomic mutation. *PRL* genes have T-DMRs. T-DMRs near CpG island-free genes are hypomethylated in somatic tissues, whereas these same regions are often methylated in ES and embryonic germ cells which are integrated into germ cell lineages (Sato et al. 2010). These data suggest that DNA methylation contributes to the evolution of *PRL* family members and their gain or loss of gene regulation.

Many gene families exist as clusters in the mammalian genome, with or without inverted repeats. Such clusters were formed by tandem gene duplication. The DNA methylation status of each T-DMR of the *PRL* superfamily members were well matched within the clusters tentatively categorized based on orientation and localization. Gene duplication may be caused by unequal crossover, which occurs most readily if a tandemly repeated gene exists, and, therefore, the rate of occurrence of gene duplication becomes high once the gene cluster is formed (Long 2001). Thus, the cluster of genes may be a unit regulated by an epigenetic mechanism, as found in the *Nanog* locus (Muramoto et al. 2010). Considering that genome stability is a factor for gene duplication by crossover, DNA methylation of gene cluster units, rather than each individual gene, may be the process for gene duplication. Alternatively, duplication of DNA elements could be a target of DNA methylation, as indicated by the study of repetitive transposable elements (Muramoto et al. 2010). Thus, gene clusters may be units of CpG methylation in the *PRL* family.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

PRL superfamily members have T-DMRs with fewer CpGs, although the number of CpGs varies among the members. Overall, the GC content and CpG frequencies are lower in the mammalian genome, suggesting that the reduction in the number of CpGs is a common event in mammalian genomic evolution. Half of the tissue-specific genes are CpG-less genes (Saxonov et al. 2006). Examples of genes with T-DMRs with fewer CpGs are mouse Sry (Nishino et al. 2004) and rat *PL-I* (Cho et al. 2001), while CpG-rich T-DMRs are found in *Oct4* as well as in genes with CpG islands (Imamura et al. 2001; Hattori et al. 2004; Suzuki et al. 2007). Recent genome-wide studies revealed that there are many T-DMRs in genes having CpG islands (Shiota et al. 2002; Khulan et al. 2006; Yagi et al. 2008). Thus, most CpGs in mammals are uniquely hypermutable T-DMRs. A possible role of epigenetics may be to modify the mutation rate by CpG-to-TpG transition due to deamination of methylcytosine, thus modifying the rates of chromosomal rearrangement (Janion 1982; Carbone et al. 2009). Our data suggest that DNA methylation has influenced the number of CpGs in the process of evolution of *PRL* superfamily members through C-to-T mutations.

Information of DNA methylation status and T-C conversion bridges the gap between protein/cDNA sequences and genome sequences. Phylogenetic analyses are useful for understanding the relationships among members of a gene family. Protein sequences and cDNA sequences, rather than genomic DNA sequences, have been used to compare gene similarities since protein/cDNA sequences are better conserved over time than genomic DNA sequences, and some mutations in a DNA sequence may have no effect on the expressed protein sequences (Wolfe et al. 1989). Due to the gaps in DNA sequence between protein/cDNA sequences and genomic sequences, we cannot simply use the phylogenic tree generated from cDNA sequences to deduce the evolutionary process of the gene families. Our trial of reconstitution of a phylogenic tree by T-C conversion, considering the possible mutation of C-T by CpG methylation, successfully produced a phylogenic tree that reflects the history of gene duplication events. The reconstitution of a phylogenic tree by T-C conversion will be useful in the case of genes with T-DMRs for understanding gene family evolution in the mammalian genome.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was funded by grants from Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT), Japan (21221008 to Kunio Shiota), and the National Institute of Biomedical Innovation (NIBIO), Japan. We thank Dr. Naoko Hattori and Dr. Shinya Sato for their helpful suggestions regarding analysis.

References

Bhattacharyya S, Lin J, Linzer D. Reactivation of a hematopoietic endocrine program of pregnancy contributes to recovery from thrombocytopenia. Mol Endocrinol. 2002; 16:1386–1393. [PubMed: 12040023]

Carbone L, Harris RA, Vessere GM, Mootnick AR, Humphray S, Rogers J, Kim SK, Wall JD, Martin D, Jurka J, Milosavljevic A, de Jong PJ. Evolutionary breakpoints in the gibbon suggest association

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

between cytosine methylation and karyotype evolution. PLoS Genet. 2009; 5:e1000538. [PubMed: 19557196]

Cho J, Kimura H, Minami T, Ohgane J, Hattori N, Tanaka S, Shiota K. DNA methylation regulates placental lactogen I gene expression. Endocrinology. 2001; 142:3389–3396. [PubMed: 11459782]

Choong ML, Tan AC, Luo B, Lodish HF. A novel role for proliferin-2 in the ex vivo expansion of hematopoietic stem cells. FEBS Lett. 2003; 550:155–162. [PubMed: 12935903]

Coulondre C, Miller JH, Farabaugh PJ, Gilbert W. Molecular basis of base substitution hotspots in *Escherichia coli*. Nature. 1978; 274:775–780. [PubMed: 355893]

Dai G, Wang D, Liu B, Kasik JW, Müller H, White RA, Hummel GS, Soares MJ. Three novel paralogs of the rodent prolactin gene family. J Endocrinol. 2000; 166:63–75. [PubMed: 10856884]

Fassett J, Nilsen-Hamilton M. Mrp3, a mitogen-regulated protein/proliferin gene expressed in wound healing and in hair follicles. Endocrinology. 2001; 142:2129–2137. [PubMed: 11316781]

Hattori N, Nishino K, Ko Y, Ohgane J, Tanaka S, Shiota K. Epigenetic control of mouse *Oct-4* gene expression in embryonic stem cells and trophoblast stem cells. J Biol Chem. 2004; 279:17063–17069. [PubMed: 14761969]

Hiraoka Y, Ogawa M, Sakai Y, Takeuchi Y, Komatsu N, Shiozawa M, Tanabe K, Aiso S. *PLP-I*: a novel prolactin-like gene in rodents. Biochim Biophys Acta. 1999; 1447:291–297. [PubMed: 10542329]

Ho Y, Liebhaber SA, Cooke NE. Activation of the human GH gene cluster: roles for targeted chromatin modification. Trends Endocrinol Metab. 2004; 15:40–45. [PubMed: 14693425]

Ho-Chen J, Bustamante J, Soares M. Prolactin-like protein-f subfamily of placental hormones/cytokines: responsiveness to maternal hypoxia. Endocrinology. 2007; 148:559–565. [PubMed: 17095594]

Imamura T, Ohgane J, Ito S, Ogawa T, Hattori N, Tanaka S, Shiota K. CpG island of rat sphingosine kinase-1 gene: tissue-dependent DNA methylation status and multiple alternative first exons. Genomics. 2001; 76:117–125. [PubMed: 11560121]

Janion C. Influence of methionine on the mutation frequency in *Salmonella typhimurium*. Mutat Res. 1982; 94:331–338. [PubMed: 7050692]

Khulan B, Thompson RF, Ye K, Fazzari MJ, Suzuki M, Stasiek E, Figueroa ME, Glass JL, Chen Q, Montagna C, Hatchwell E, Selzer RR, Richmond TA, Green RD, Melnick A, Greally JM. Comparative isoschizomer profiling of cytosine methylation: the HELP assay. Genome Res. 2006; 16:1046–1055. [PubMed: 16809668]

Kimura F, Takakura K, Takebayashi K, Ishikawa H, Kasahara K, Goto S, Noda Y. Messenger ribonucleic acid for the mouse decidual prolactin is present and induced during in vitro decidualization of endometrial stromal cells. Gynecol Endocrinol. 2001; 15:426–432. [PubMed: 11826766]

Kimura AP, Sizova D, Handwerger S, Cooke NE, Liebhaber SA. Epigenetic activation of the human growth hormone gene cluster during placental cytotrophoblast differentiation. Mol Cell Biol. 2007; 27:6555–6568. [PubMed: 17636034]

Li E. Chromatin modification and epigenetic reprogramming in mammalian development. Nat Rev Genet. 2002; 3:662–673. [PubMed: 12209141]

Lin J, Poole J, Linzer DI. Three new members of the mouse prolactin/growth hormone family are homologous to proteins expressed in the rat. Endocrinology. 1997; 138:5541–5549. [PubMed: 9389542]

Linzer DI, Nathans D. Nucleotide sequence of a growth-related mRNA encoding a member of the prolactin-growth hormone family. Proc Natl Acad Sci USA. 1984; 81:4255–4259. [PubMed: 6087314]

Long M. Evolution of novel genes. Curr Opin Genet Dev. 2001; 11:673–680. [PubMed: 11682312]

Louis E. Evolutionary genetics: making the most of redundancy. Nature. 2007; 449:673–674. [PubMed: 17928851]

Mallon AM, Wilming L, Weekes J, Gilbert JG, Ashurst J, Peyrefitte S, Matthews L, Cadman M, McKeone R, Sellick CA, Arkell R, Botcherby MR, Strivens MA, Campbell RD, Gregory S, Denny P, Hancock JM, Rogers J, Brown SD. Organization and evolution of a gene-rich region of the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

mouse genome: a 12.7-Mb region deleted in the Del(13)Svea36H mouse. Genome Res. 2004; 14:1888–1901. [PubMed: 15364904]

Müller H, Ishimura R, Orwig KE, Liu B, Soares MJ. Homologues for prolactin-like proteins A and B are present in the mouse. Biol Reprod. 1998; 58:45–51. [PubMed: 9472921]

Muramoto H, Yagi S, Hirabayashi K, Sato S, Ohgane J, Tanaka S, Shiota K. Enrichment of short interspersed transposable elements to embryonic stem cell-specific hypomethylated gene regions. Genes Cells. 2010; 15:855–865. [PubMed: 20629982]

Nishino K, Hattori N, Tanaka S, Shiota K. DNA methylation-mediated control of *Sry* gene expression in mouse gonadal development. J Biol Chem. 2004; 279:22306–22313. [PubMed: 14978045]

Orwig KE, Ishimura R, Müller H, Liu B, Soares MJ. Identification and characterization of a mouse homolog for decidual/trophoblast prolactin-related protein. Endocrinology. 1997; 138:5511–5517. [PubMed: 9389538]

Rakyan V, Down T, Thorne N, Flicek P, Kulesha E, Gräf S, Tomazou E, Bäckdahl L, Johnson N, Herberth M, Howe K, Jackson D, Miretti M, Fiegler H, Marioni J, Birney E, Hubbard T, Carter N, Tavaré S, Beck S. An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). Genome Res. 2008; 18:1518–1529. [PubMed: 18577705]

Russell GJ, Walker PM, Elton RA, Subak-Sharpe JH. Doublet frequency analysis of fractionated vertebrate nuclear DNA. J Mol Biol. 1976; 108:1–23. [PubMed: 1003479]

Sato S, Yagi S, Arai Y, Hirabayashi K, Hattori N, Iwatani M, Okita K, Ohgane J, Tanaka S, Wakayama T, Yamanaka S, Shiota K. Genome-wide DNA methylation profile of tissue-dependent and differentially methylated regions (T-DMRs) residing in mouse pluripotent stem cells. Genes Cells. 2010; 15:607–618. [PubMed: 20477876]

Saxonov S, Berg P, Brutlag DL. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proc Natl Acad Sci USA. 2006; 103:1412–1417. [PubMed: 16432200]

Shiota K. DNA methylation profiles of CpG islands for cellular differentiation and development in mammals. Cytogenet Genome Res. 2004; 105:325–334. [PubMed: 15237220]

Shiota K, Kogo Y, Ohgane J, Imamura T, Urano A, Nishino K, Tanaka S, Hattori N. Epigenetic marks by DNA methylation specific to stem, germ and somatic cells in mice. Genes Cells. 2002; 7:961–969. [PubMed: 12296826]

Simmons D, Rawn S, Davies A, Hughes M, Cross J. Spatial and temporal expression of the 23 murine prolactin/placental lactogen-related genes is not associated with their position in the locus. BMC Genomics. 2008; 9:352. [PubMed: 18662396]

Soares M. The prolactin and growth hormone families: pregnancy-specific hormones/cytokines at the maternal-fetal interface. Reprod Biol Endocrinol. 2004; 2:51. [PubMed: 15236651]

Soares M, Alam S, Duckworth M, Horseman N, Konno T, Linzer D, Maltais L, Nilsen-Hamilton M, Shiota K, Smith J, Wallis M. A standardized nomenclature for the mouse and rat prolactin superfamilies. Mamm Genome. 2007a; 18:154–156. [PubMed: 17476555]

Soares MJ, Konno T, Alam SM. The prolactin family: effectors of pregnancy-dependent adaptations. Trends Endocrinol Metab. 2007b; 18:114–121. [PubMed: 17324580]

Suzuki M, Sato S, Arai Y, Shinohara T, Tanaka S, Greally JM, Hattori N, Shiota K. A new class of tissue-specifically methylated regions involving entire CpG islands in the mouse. Genes Cells. 2007; 12:1305–1314. [PubMed: 18076568]

Thompson J, Higgins D, Gibson T. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994; 22:4673–4680. [PubMed: 7984417]

Toft DJ, Linzer DI. Prolactin (PRL)-like protein J, a novel member of the PRL/growth hormone family, is exclusively expressed in maternal decidua. Endocrinology. 1999; 140:5095–5101. [PubMed: 10537137]

Toft DJ, Rosenberg SB, Bergers G, Volpert O, Linzer DI. Reactivation of proliferin gene expression is associated with increased angiogenesis in a cell culture model of fibrosarcoma tumor progression. Proc Natl Acad Sci USA. 2001; 98:13055–13059. [PubMed: 11606769]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Tomikawa J, Fukatsu K, Tanaka S, Shiota K. DNA methylation-dependent epigenetic regulation of dimethylarginine dimethylaminohydrolase 2 gene in trophoblast cell lineage. J Biol Chem. 2006; 281:12163–12169. [PubMed: 16520373]

Wiemers D, Shao L, Ain R, Dai G, Soares M. The mouse prolactin gene family locus. Endocrinology. 2003; 144:313–325. [PubMed: 12488360]

Wolfe KH, Sharp PM, Li WH. Mutation rates differ among regions of the mammalian genome. Nature. 1989; 337:283–285. [PubMed: 2911369]

Yagi S, Hirabayashi K, Sato S, Li W, Takahashi Y, Hirakawa T, Wu G, Hattori N, Ohgane J, Tanaka S, Liu XS, Shiota K. DNA methylation profile of tissue-dependent and differentially methylated regions (T-DMRs) in mouse promoter regions demonstrating tissue-specific gene expression. Genome Res. 2008; 18:1969–1978. [PubMed: 18971312]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

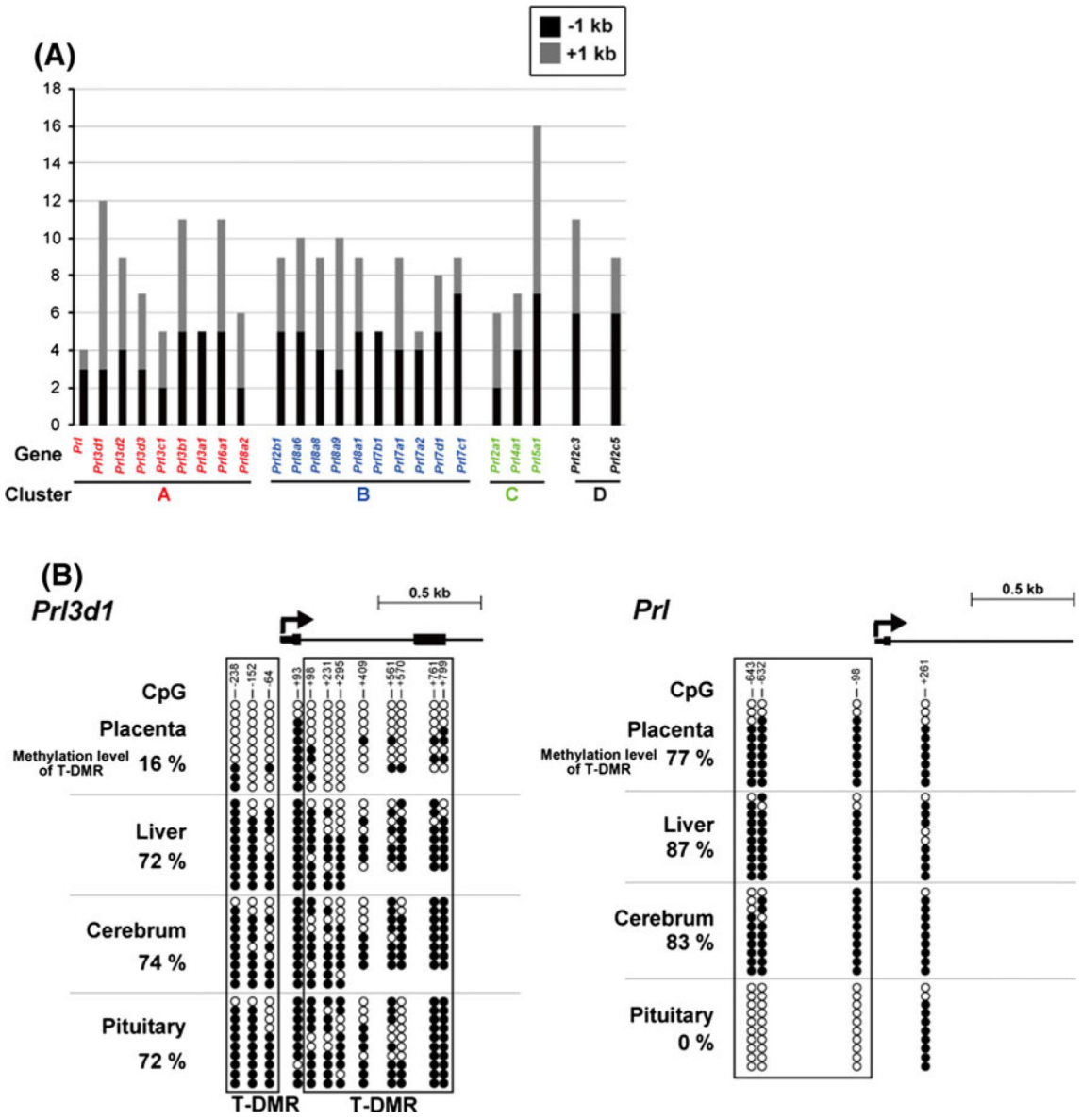

Fig. 1.

DNA methylation status of *Prl* and *Prl3d1* in the placenta, cerebrum, liver, and pituitary. **a** The number of CpGs located within ±1 kb from TSS of *PRL* family genes. *Black* and *gray bars* indicate the number of CpG in −1 and +1 kb, respectively, from TSS in *PRL* members. **b** DNA methylation status of CpGs located from −1 kb to +1 kb in *Prl3d1* and *Prl*, as determined by bisulfite sequencing. *Vertical lines* and *numbers* indicate the position of cytosine residues of CpGs relative to the transcription start site (+1). *Open dots* and *filled dots* indicate unmethylated and methylated CpGs, respectively. The *open squares* around CpGs indicate T-DMRs

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

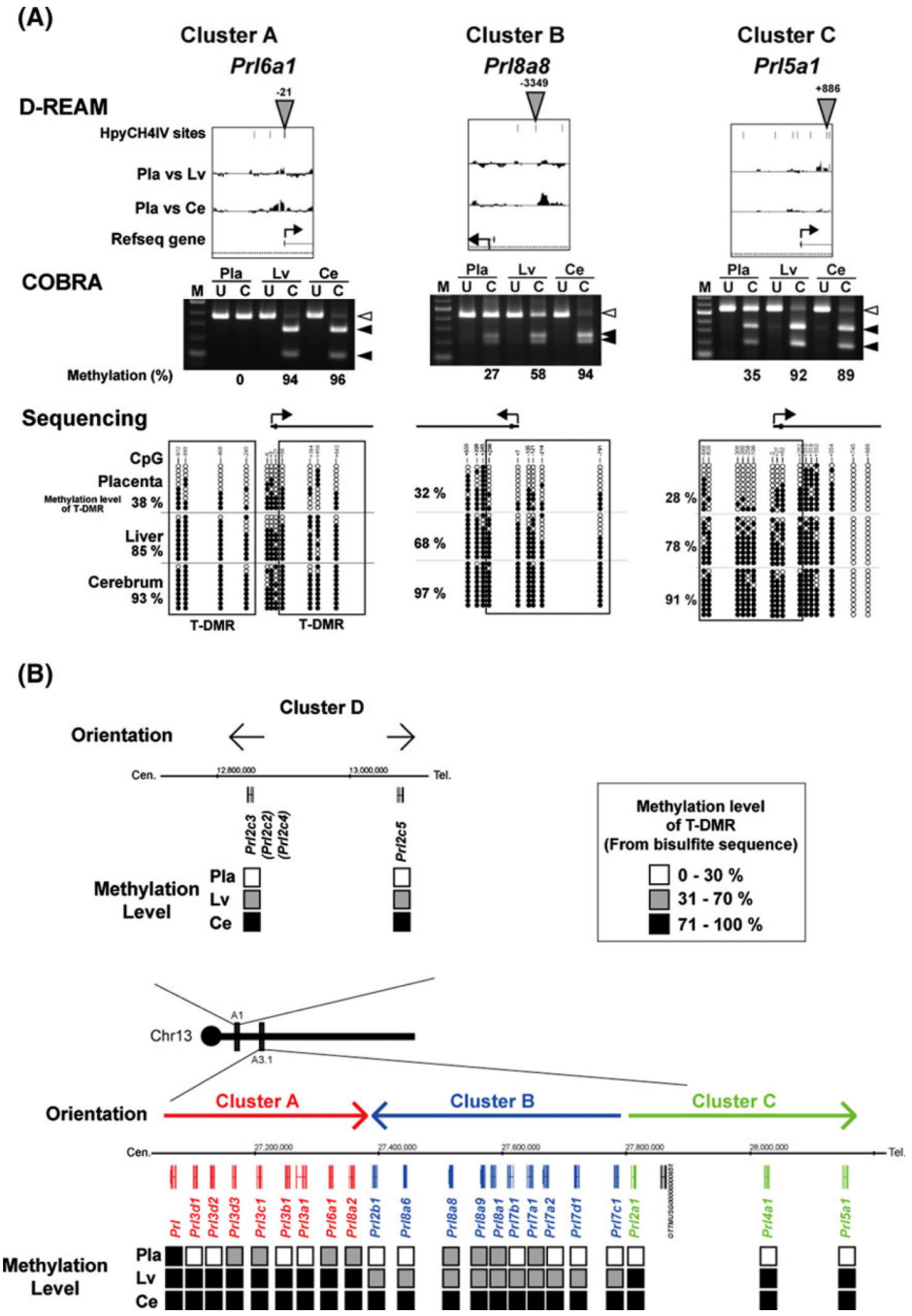

Fig. 2.

DNA methylation status of *PRL* superfamily members in the placenta, liver, and cerebrum. **a** The upper panels show the integrated genome browser of comparative D-REAM signals. *Arrowheads* and *vertical lines* represent T-DMR and HpyCH4IV sites, respectively. Comparative signals between tissues are indicated as *bars*. *Pla* Placenta, *Lv* liver, *Ce* cerebrum. The central panel shows electrophoresis patterns for COBRA using HpyCH4IV. Where PCR products are fragmented (*filled arrowhead*), the HpyCH4IV sites are methylated, and where PCR products are not fragmented (*open arrowhead*), the HpyCH4IV

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

sites are unmethylated. The level of methylation, estimated by the intensity of each band, is indicated below the gels. Undigested (*U*), digested (*C*), and 100-bp DNA ladder (*M*) are indicated. The lower panels show the DNA methylation status of the CpGs within the 2 kb flanking the transcription start sites in *PRL* superfamily genes in the placenta, liver, and cerebrum, investigated by bisulfite sequencing. The *vertical lines* and *numbers* indicate the position of cytosine residues of CpGs relative to the transcription start site (+1). CpGs are represented as *open dots* (unmethylated) or *filled dots* (methylated). The *open square* around the CpGs indicates T-DMR. **b** Correlation between the DNA methylation profile of the *PRL* gene superfamily and gene orientation and location. According the genomic location and orientation, *PRL* superfamily genes were categorized into four clusters: cluster A (*red*), cluster B (*blue*), cluster C (*green*), and cluster D (*black*). *Open, gray,* and *black boxes* indicate DNA methylation levels of T-DMRs from the analysis of bisulfite sequencing (0–30%, 31–70%, and 71–100%, respectively). *Cen* Centromere, *Tel* telomere, *Pla* placenta, *Lv* liver, *Ce* cerebrum

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

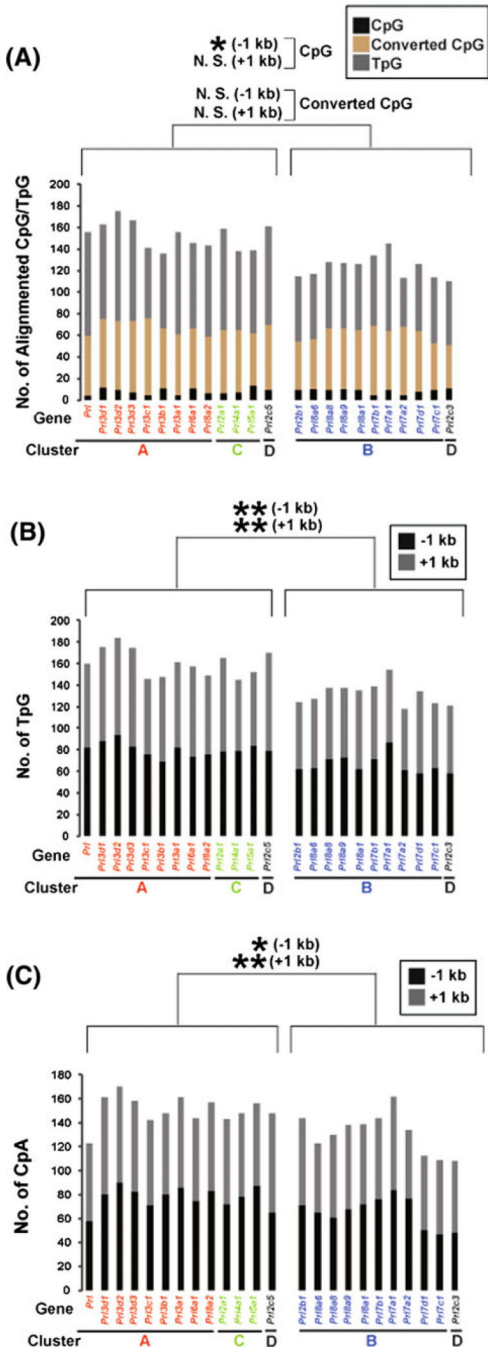

Fig. 3.

Correlation between the DNA methylation profile of *PRL* superfamily genes and the number of CpGs, TpGs, and CpAs in a 2 kb region flanking the TSS. **a** The number of CpGs and converted CpGs located within ±1 kb from the TSS of *PRL* family genes. For the TpG-CpG-converted phylogenic tree, conserved regions were selected by sequence alignment. Then, TpGs in the conserved sequence were converted to CpGs. *Black bar*, the number of CpGs located within ±1 kb from the TSS; *brown bar*, that of converted CpGs; *gray bar*, that of TpGs. **P* < 0.05; ***P* < 0.01; *N.S.* Not significant (Wilcoxon signed-rank test). **b** The

number of TpGs located within ±1 kb from the TSS of *PRL* family genes. **c** The number of CpAs located within ±1 kb from the TSS of *PRL* family genes

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

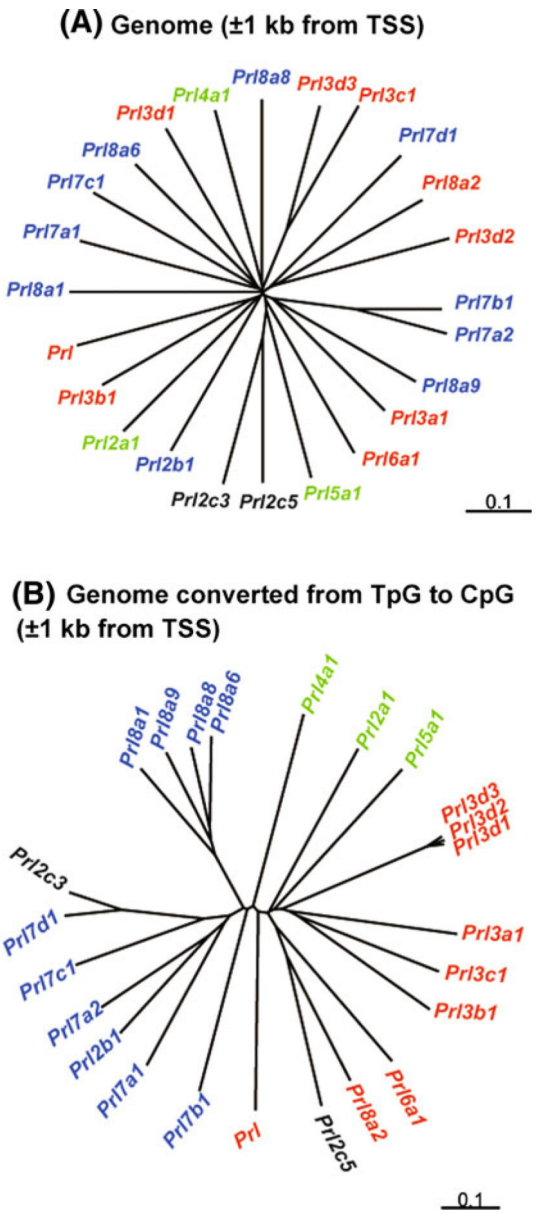

Fig. 4.

Correlation between the DNA methylation profile of *PRL* superfamily members and sequence properties. **a** The relationship between genomic sequence (±1 kb of the TSS) and DNA methylation status. Comparisons of the 24 paralogous mouse *PRL* genomic sequences and the phylogenetic tree construction were performed using the CLUSTAL W software program. **b** The relationship between genomic sequences converted from CpG to TpG and DNA methylation status. The TpG sequence located from −1 to +1 kb of TSS was converted to CpG (Fig. 3a). These sequences were used to construct to a phylogenetic tree

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1

Genomic location and expression patterns of mouse PRL superfamily members

Genomic location	Cluster	Official symbol	Official alias symbol	Orientation[a]	Expression[b]		
					Pituitary	Placenta	Others
Chr 13 A3.1	A	*Prl*	Prl1a1	+	Y	N	Y[c]
		Prl3d1	PL-Iα	+	N	Y	N
		Prl3d2	PL-Iβ	+	N	Y	N
		Prl3d3	PL-Iγ	+	N	Y	N
		Prl3c1	PLP-J	+	N	N	Y[c]
		Prl3b1	PL-II	+	N	Y	N
		Prl3a1	PLP-I	+	N	Y	N
		Prl6a1	PLP-B	+	N	Y	Y[c]
		Prl8a2	dPRP	+	N	Y	Y[c]
	B	*Prl2b1*	PLP-K	–	N	Y	N
		Prl8a6	PLP-Cα	–	N	Y	N
		Prl8a8	PLP-Cγ	–	N	Y	N
		Prl8a9	PLP-Cβ	–	N	Y	N
		Prl8a1	PLP-Cδ	–	N	Y	N
		Prl7b1	PLP-N	–	N	Y	N
		Prl7a1	PLP-E	–	N	Y	Y[d]
		Prl7a2	PLP-F	–	N	Y	N
		Prl7d1	PLF-RP	–	N	Y	N
		Prl7c1	PLP-O	–	N	Y	N
	C	*Prl2a1*	PLP-M	+	N	Y	N
		Prl4a1	PLP-A	+	N	Y	N
		Prl5a1	PLP-L	+	N	Y	N
Chr13 A1	D	*Prl2c3*	PLF2	–	N	N	Y[e]
		Prl2c5	MRP4	+	N	N	Y[f]

[a] + the gene transcribed from centromere to telomere, – the gene transcribed from telomere to centromere

[b] *Y* express, *N* nonexpress (Lin et al. 1997; Orwig et al. 1997; Müller et al. 1998; Hiraoka et al. 1999; Toft and Linzer 1999; Dai et al. 2000; Fassett and Nilsen-Hamilton 2001; Kimura et al. 2001; Toft et al. 2001; Bhattacharyya et al. 2002; Choong et al. 2003; Ho-Chen et al. 2007)

[c] Decidua

[d] Hematopoietic cells

[e] Fibroblast/fibrosarcoma

[f] Fibroblast