

Research article

Open Access

Functional genomics of HMGN3a and SMARCAL1 in early mammalian embryogenesis

Alper Uzun^{†1}, Nelida Rodriguez-Osorio^{†2,3}, Abdullah Kaya⁴,
Hongfeng Wang², John J Parrish⁴, Valentin A Ilyin^{*1} and Erdogan Memili^{*2}

Address: ¹Department of Biology, Northeastern University, Boston, MA, USA, ²Department of Animal and Dairy Science, Mississippi State University, Mississippi State, MS, USA, ³Grupo Centauro Universidad de Antioquia, Medellín, Antioquia, Colombia and ⁴Department of Animal Sciences, University of Wisconsin-Madison, Madison, WI, USA

Email: Alper Uzun - uzun.a@neu.edu; Nelida Rodriguez-Osorio - nerodriguez@agronica.udea.edu.co; Abdullah Kaya - akaya@wisc.edu; Hongfeng Wang - hw179@ads.msstate.edu; John J Parrish - parrish@ansci.wisc.edu; Valentin A Ilyin* - v.ilyin@neu.edu; Erdogan Memili* - em149@ads.msstate.edu

* Corresponding authors †Equal contributors

Published: 24 April 2009

Received: 3 July 2008

BMC Genomics 2009, **10**:183 doi:10.1186/1471-2164-10-183

Accepted: 24 April 2009

This article is available from: <http://www.biomedcentral.com/1471-2164/10/183>

© 2009 Uzun et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Embryonic genome activation (EGA) is a critical event for the preimplantation embryo, which is manifested by changes in chromatin structure, transcriptional machinery, expression of embryonic genes, and degradation of maternal transcripts. The objectives of this study were to determine transcript abundance of HMGN3a and SMARCAL1 in mature bovine oocytes and early bovine embryos, to perform comparative functional genomics analysis of these genes across mammals.

Results: New annotations of both HMGN3a and SMARCAL1 were submitted to the Bovine Genome Annotation Submission Database at BovineGenome.org. Careful analysis of the bovine SMARCAL1 consensus gene set for this protein (GLEAN_20241) showed that the NCBI protein contains sequencing errors, and that the actual bovine protein has a high degree of homology to the human protein. Our results showed that there was a high degree of structural conservation of HMGN3a and SMARCAL1 in the mammalian species studied. HMGN3a transcripts were present at similar levels in bovine matured oocytes and 2–4-cell embryos but at higher levels in 8–16-cell embryos, morulae and blastocysts. On the other hand, transcript levels of SMARCAL1 decreased throughout preimplantation development.

Conclusion: The high levels of structural conservation of these proteins highlight the importance of chromatin remodeling in the regulation of gene expression, particularly during early mammalian embryonic development. The greater similarities of human and bovine HMGN3a and SMARCAL1 proteins may suggest the cow as a valuable model to study chromatin remodeling at the onset of mammalian development. Understanding the roles of chromatin remodeling proteins during embryonic development emphasizes the importance of epigenetics and could shed light on the underlying mechanisms of early mammalian development.

Background

Early embryonic development is initiated when mature oocytes (MII) are fertilized by spermatozoa. Maternal factors, such as mRNAs, microRNAs and proteins stored in the oocyte, provide the means of support for the first few days of development. The transition from a maternal to a zygotic control of development, called maternal to zygotic transition (MZT), and the activation of the embryonic genome involve chromatin structural modifications that take place during the first few embryonic cell cycles [1]. Embryonic genome activation (EGA) sets the stage for later development [2,3]. Changes in chromatin structure have been characterized throughout the transition from transcriptional incompetence to the minor activation of the zygotic genome at the 1-cell stage and through the major genome activation at the 2-cell stage in murine embryos [4]. In bovine embryos EGA occurs at the 8- to 16-cell stage with extensive programming of gene expression. However, the regulation of chromatin remodeling during EGA still remains a mystery.

Chromatin remodeling is an extensive process occurring during early embryogenesis. An essential property of the embryonic chromatin structure is to prevent the access of the transcriptional machinery to all of the promoters in the genome. The expression of some genes may be mediated by chromatin remodeling proteins. Chromatin remodeling complexes may change the overall pattern of expression of mammalian genes, allowing transcription factors and signaling pathways to produce different genomic transcriptional responses to common signals [5]. This is particularly important for preimplantation embryos starting cell differentiation cascades that will lead to tissue and organogenesis. These changes in chromatin structure generate activation of the transcriptional machinery and gene expression occurring during early embryo development, leading to a unique chromatin structure capable of maintaining totipotency during embryogenesis and differentiation during postimplantation development [3].

The High Mobility Group Nucleosomal (HMGN) protein family is the only group of nuclear proteins that bind to the 147-base pair long nucleosome core particle with no sequence specificity [6]. HMGN proteins are present in the nuclei of all mammalian and most vertebrate cells at approximately 10% of the abundance of histones [7]. They bind as homodimers to the nucleosome and cause chromatin modifications that facilitate and enhance several DNA-dependent activities, such as transcription, replication and DNA repair. This protein family is composed of 3 members, HMGN1 (also known as HMG-14), HMGN2 (also known as HMG-17), and the most recently discovered HMGN3, initially named TRIP7 for its ability to bind the thyroid hormone receptor [8].

In the mouse HMGN1 and HMGN2 have been detected throughout oogenesis and preimplantation development and are progressively down-regulated throughout the entire embryo, except in cell types undergoing active differentiation [9]. Reduction in the levels of HMGN1 and 2 mRNA also occurs during myogenesis in rat, suggesting that down-regulation of HMGN mRNA may be associated with tissue differentiation [10]. Depletion of HMGN1 and HMGN2 in one- or two-cell embryos delays subsequent embryonic divisions. Cells derived from HMGN1^{-/-} mice have an altered transcription profile and are hypersensitive to stress [9]. Experimental manipulations of the intracellular levels of HMGN1 in *X. laevis* embryos cause specific developmental defects at the post-blastula stages. Furthermore, HMGN proteins regulate the expression of specific genes during *X. laevis* development [11]. Several lines of evidence implicate HMGN1 and 2 in transcriptional regulation. Chromatin containing genes that are actively being transcribed has two- to three times more HMGN1 and 2 compared with total chromatin [9].

The human HMGN3 transcript produces two splice variants HMGN3a the long isoform with 99 amino acids, and HMGN3b with 77 amino acids that arises due to a truncation of the fifth exon. Although no HMGN3b protein has been identified in the rat and cow, ESTs with high identity to it suggest that this splice variant may also exist in these species. The cow, mouse, and rat HMGN3a proteins share more than 81% identity with the human HMGN3a protein [8]. The role of HMGN3a has not been studied in mammalian development. Our previous data show that HMGN3a is expressed at similar levels in oocytes and 8-cell bovine embryos [3]. We have detected high HMGN3a mRNA levels in IVF produced bovine blastocysts. Furthermore, HMGN3a was significantly higher in IVF derived blastocysts compared to blastocysts produced by somatic cell chromatin transfer (SCCT), which had lower levels of HMGN3a transcript similar to those detected in somatic cells (unpublished data). Although the exact function of HMGN3a during early embryonic development has not been determined, its role in facilitating chromatin modifications and enhancing transcription, replication, and DNA repair is critical for early embryo development [8].

Another important mechanism in regulation of chromatin structure in the early embryo is mediated by nucleosome repositioning factors, which are ATP-dependent chromatin-remodeling enzymes. Nucleosome repositioning factors use energy released by ATP hydrolysis to alter histone-DNA contacts and reposition nucleosomes to create chromatin environments that are either open or compact. These factors do not involve sequence specific DNA binding sites, but rather are recruited onto promoter regions by specific transcription factors. Nucleosome repositioning factors typically exist as multi subunit pro-

tein complexes, like the SWI/SNF (from SWItching and Sucrose Non-Fermenting in yeast) ATP-dependent chromatin remodeling complex [12]. SWI/SNF complexes are thought to regulate transcription of certain genes by altering the chromatin structure around them with their helicase and ATPase activities [13,14]. In mammals, each SWI/SNF complex has any of two distinct ATPases as the catalytic subunit of SMARCA2 (also known as BRM or Brahma) and SMARCA4 (also known as BRG1 Brahma related gene 1) [15]. Both ATPases have important developmental functions. In primates, expression of both subunits remains constant and low throughout embryogenesis until the blastocyst stage [16]. In mouse embryos, Smarca4 transcripts remain at stable levels throughout preimplantation development, while Smarca2 transcripts remain low until the blastocyst stage, when its mRNA levels increase [17]. In porcine embryos, SMARCA2 transcripts are most abundant in germinal vesicle (GV) stage oocytes and decline progressively during embryo development to blastocyst stage [18]. Mutant mice lacking the Smarca4 gene die at preimplantation while the Smarca2-null mouse mutant is viable and shows a mild overgrowth phenotype [19,20].

Another member of the SWI/SNF family of proteins involved in chromatin remodeling is SMARCA1 (SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a member 1), considered a global transcription activator and also called SNF2L1. Like other SWI/SNF members, the SMARCA1 protein has a helicase ATP-binding domain. However, since the rest of its motifs diverge from other members of the SWI/SNF family, it has been classified in the ISWI (for Imitation SWItch) subfamily of ATPases, together with SMARCA5. Decreasing levels of SMARCA5 were found during Rhesus monkey embryogenesis from GV oocytes until blastocyst stage. The same study reported low levels of SMARCA1 throughout all stages of embryogenesis except for the 8-cell stage [16].

Members of the SNF2 subfamily of SWI/SNF proteins are characterized by its seven motifs (I, Ia, II, III, IV, V and VI) [21]. SMARCAL1 (SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a-like 1) is one of the SNF2 members and shows high sequence similarity to the *E. coli* RNA polymerase-binding protein HepA [21]. Recent reports have linked mutations in the SMARCAL1 gene with Schimke immunosseous dysplasia (SIOD), a human autosomal recessive disorder with the diagnostic features of spondyloepiphyseal dysplasia, renal dysfunction, and T-cell immunodeficiency [22]. The ability of SMARCAL1, to interact primarily with nucleosomes was demonstrated using protein interaction microarrays. SMARCAL1 transcripts are ubiquitously expressed in different human and mouse tissues, suggesting a role in

normal cellular functions or housekeeping activities, such as transcriptional regulation [21]. Although no studies have reported the expression of SMARCAL1 during early embryogenesis in mammals, we previously detected a 7-fold increase of the SMARCAL1 mRNA in 8-cell embryos as compared with MII oocytes by using oligonucleotide microarray gene expression analysis and Real Time PCR validation [3].

Additionally, studies on the SWI/SNF complex associated factor SMARCC1 (also called SRG3 and BAF155), a core subunit of the SWI/SNF complex, have highlighted the importance of the ATPase subunits and the whole complex during embryogenesis. In the absence of Smarcc1, mouse embryonic development ceased during preimplantation stages, indicating that Smarcc1, as well as the chromatin-remodeling process, plays an essential role in early mouse development [23]. SMARCC1 mRNA was found in high levels in GV stage Rhesus monkey oocytes and at very low levels throughout early embryogenesis but was higher later at the hatched blastocyst stage [16].

The limited availability of fully annotated bovine genes has been a limitation for bovine genomic studies. Many bovine proteins are only partially annotated or are based on computational prediction. The objectives of this study were to determine transcript abundance of HMGN3a and SMARCAL1 in mature bovine oocytes and early bovine embryos, to perform comparative functional genomics analysis of these genes across mammals, including humans, annotate and analyze the conserved/non conserved regions of them on the comparative modeled structure.

Results and discussion

HMGN3a and SMARCAL1 transcripts are highly expressed in bovine oocytes and early embryos

The mRNA isolated from oocytes and early embryos exhibited typical ribosomal RNA band ratio (28S:18S). Blastocysts showed a ratio of 28S:18S bands closer to 2.0, similar to somatic cells. Using real time PCR, we demonstrated the presence of HMGN3a in bovine MII oocytes, 2–4-cell embryos, 8–16-cell embryos, morulae and blastocysts. HMGN3a transcript abundance was significantly lower in MII oocytes and 2-cell embryos compared to 8-cell, morula and blastocyst stage embryos (Figure 1A). SMARCAL1 transcript abundance was similar in MII oocytes and 8-cell embryos. However we observed significantly higher levels of SMARCAL1 mRNA in 2-cell embryos. SMARCAL1 transcripts decrease significantly in the morula and blastocyst stages (Figure 1B). This result differs from our previous microarray and Real-Time PCR gene expression analysis, which showed that SMARCAL1 was expressed at significantly higher levels in 8-cell embryos compared to bovine oocytes [3]. This difference

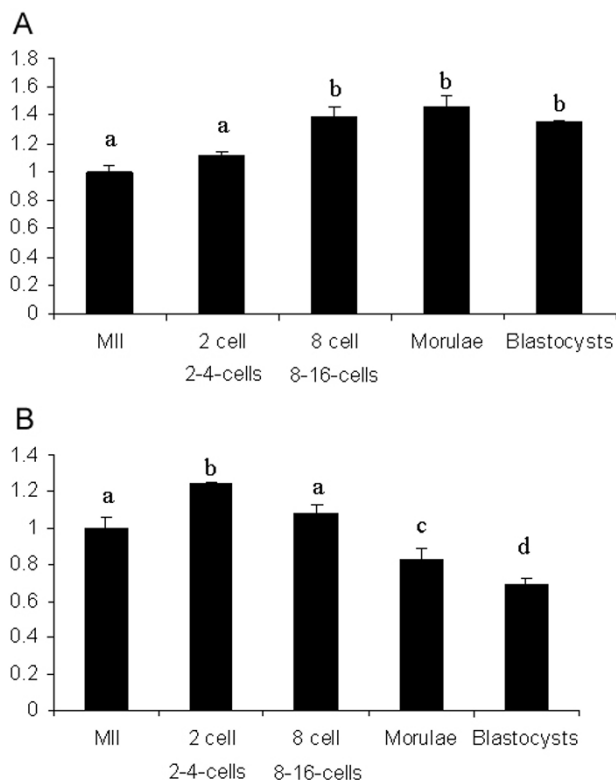


Figure 1

A. Real Time PCR analysis of HMGN3a. The mRNA abundance of HMGN3a was determined in bovine matured oocytes, 2–4-cell, 8–16-cell, morula and blastocyst stage embryos. Bars represent relative expression values of 2–4-cell embryos, 8–16-cell embryos, and blastocysts to the expression in MII oocytes. Transcript abundance was normalized to the housekeeping gene, GAPDH. Expression ratio of HMGN3 gene in the 8–16-cell, morulae and blastocysts were significant higher than that during MII and 2–4-cell stages. Different letters represent statistically significant differences ($P < 0.05$). **B. Real Time PCR analysis of SMARCAL1.** mRNA abundance SMARCAL1 was determined in bovine matured oocytes, 2–4-cell, 8–16-cell, morula and blastocyst stage embryos. Bars represent relative expression values of 2–4-cell embryos, 8–16-cell embryos, and blastocysts to the expression in MII oocytes. Transcript abundance was normalized to the housekeeping gene, GAPDH. SMARCAL1 relative expression was similar in MII oocytes and 8–16-cell embryos. However, significantly higher levels of SMARCAL1 in 2–4-cell embryos were observed. SMARCAL1 transcripts decreased significantly in the morula and blastocyst stages. Different letters represent statistically significant differences ($P < 0.05$).

might be due to the fact that in the current study we have pooled 8- and 16-cell embryos where expression of SMARCAL1 transcripts might be decreasing at the 16-cell stage. The present results confirm the expression of HMGN3a and SMARCAL1 in the bovine oocyte and early embryo. Their confirmed role in chromatin remodeling in

other tissues, could suggest involvement of these proteins in the specialized chromatin remodeling process occurring during embryo development.

Smarcc1 (SRG3) expression during mouse oogenesis and preimplantation stages was studied using immunofluorescence and western blot assays. Smarcc1 was present in the nuclei of oocytes during growth and maturation. Following fertilization, Smarcc1 was detected in higher levels in the male pronucleus compared to the female pronucleus. Expression of Smarcc1 was accompanied by expression of Smarca4 and Ini1, other core subunit of the SWI/SNF complex. The expression of these chromatin remodeling factors could suggest a role for remodeling factors in chromatin structure and function during early development [23]. These findings suggest that although the ISWI proteins are widely expressed and play important roles in promoting cellular proliferation and differentiation, they may not play a prominent role during blastocyst formation and may only become key factors during postimplantation life [16].

Comparative functional genomics analyses of HMGN3a across mammals

Seven mammalian species were used in the construction of a HMGN3a phylogenetic tree (Figure 2A). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) is shown next to the branches. All positions containing gaps and missing data were eliminated from the dataset. There were a total of 95 positions in the final dataset, of which 18 were parsimony informative. The most significant observation in multiple sequence alignment of HMGN3a was the insertion of alanine, in the fifth exon of the *Bos taurus* protein, (highlighted in red on Figure 2B). Several substitutions in the bovine sequence were shared by other mammals in the alignment.

HMGN3a constitutes a family of relatively low molecular weight non-histone components of about 100 amino acid residues. *Macaca mulatta* and *Canis familiaris* HMGN3a proteins have longer sequences with regions not shared with the other species. We focused on the regions of the protein shared by all species. Also we showed other alanine substitutions in the alignment (marked with stars) (Figure 2B).

Comparative functional genomics analyses of SMARCAL1 across mammals

SMARCAL1 has four conserved domains (Figure 3). The first and the second are two HARP (HepA-related protein) domains of approximately 60 residues long, with single-stranded DNA-dependent ATPase activity. The third conserved domain is a helicase like domain named SNF2 N-terminal domain and the fourth is a helicase C-terminal domain [21].

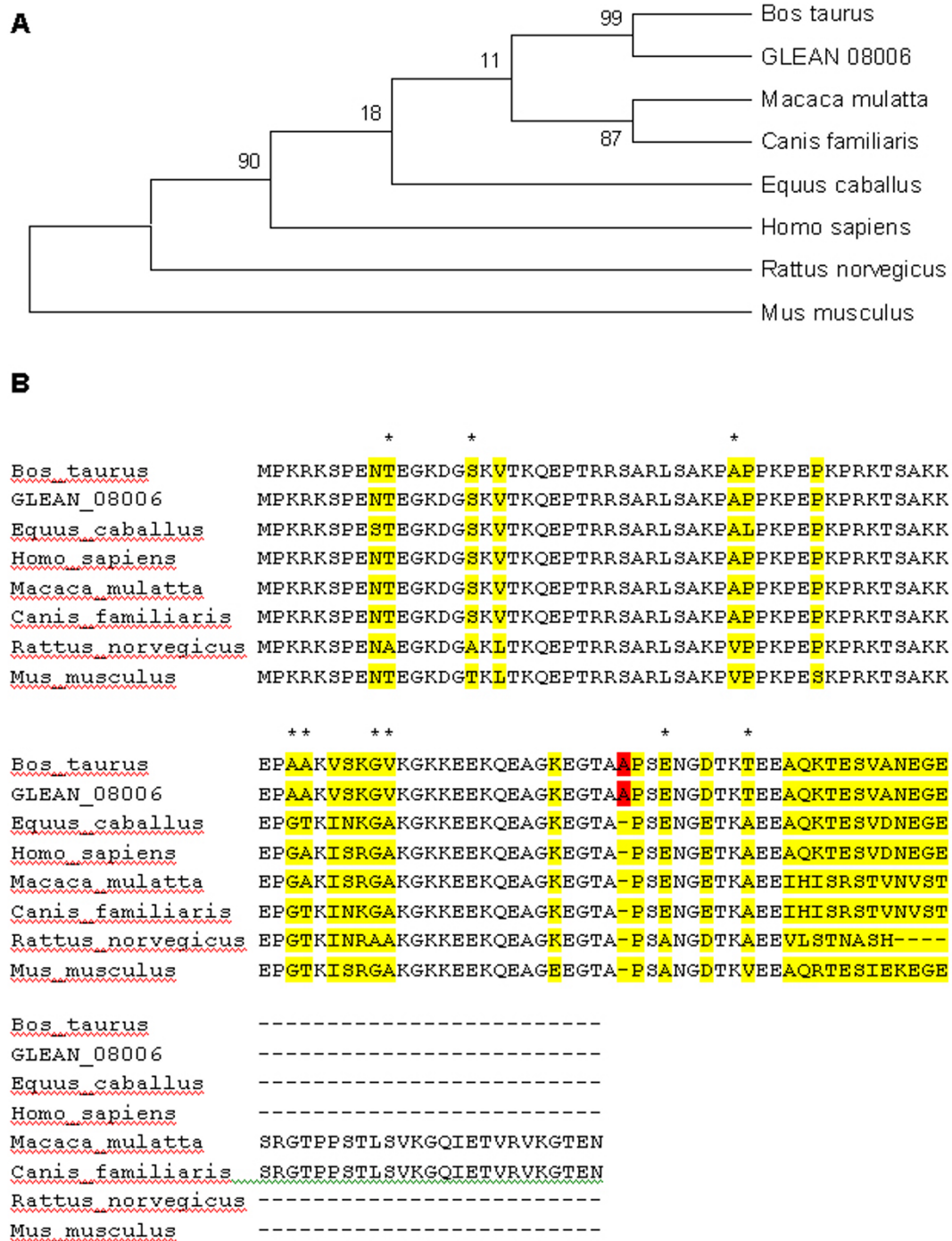


Figure 2
A. Phylogenetic tree of evolutionary relationships of HMGN3a in 7 mammalian taxa using the Maximum Parsimony method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) is shown next to the branches. All positions containing gaps and missing data were eliminated from the dataset. **B.** Multiple sequence alignment of HMGN3a. Highlighted regions show substitutions in at least 1 of the 7 species. The alignment includes both the official bovine HMGN3a gene model (GLEAN_08006), in blue, and the bovine NCBI HMGN3a protein (NP_001029676.1). The insertion of alanine in the fifth exon of the bovine protein is marked in red.

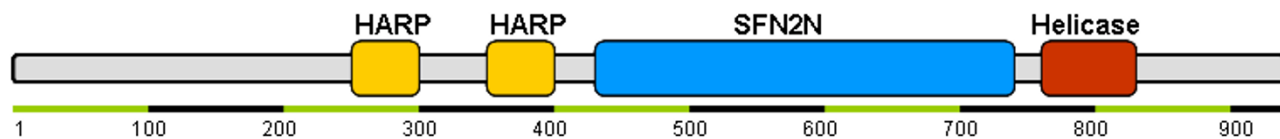


Figure 3
SMARCAL1 has 4 distinctive domains. The starting and ending residue numbers are 245–299, 342–396, 437–727, and 741–818.

We used the whole SMARCAL1 protein sequences from 9 mammalian species to construct the phylogenetic tree (Figure 4). Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap analysis (500 replicates) is shown next to the branches. All positions containing gaps and missing data were eliminated from the dataset. There were a total of 856 positions in the final dataset, of which 198 were parsimony informative. We used each one of the four SMARCAL1 conserved domains to build separate multiple sequence alignments and construct separate phylogenetic trees for each domain (Figures 5A, Six part A, Seven part A and Nine part A). Phylogenetic analysis shows that while *Homo sapiens*, *Pan troglodytes*, *Macaca mulatta* are clustering together, *Equus caballus*, *Canis familiaris* and *Bos taurus* have relatively distant position in the tree. *Rattus norvegicus* and *Mus musculus* separated these organisms in the tree. *Monodelphis domestica* becomes the most distant species among 9 mammals in the tree.

For the first HARP domain, the positions at which substitutions occur are highlighted in yellow (Figure 5B) *Mono-*

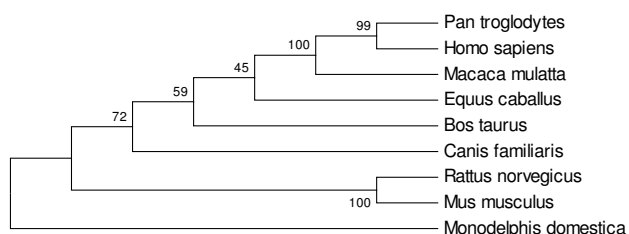


Figure 4
Phylogenetic tree of evolutionary relationships of the complete SMARCAL1 protein in 9 mammalian taxa, using the Maximum Parsimony method. The bootstrap consensus tree inferred from 500 replicates is taken to represent the evolutionary history of the taxa analyzed. All positions containing gaps and missing data were eliminated from the dataset. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) is shown next to the branches.

delphis domestica was the most distantly related mammal with respect to this domain. Substitutions were observed in 24 positions. On the 4th substitution, glutamate, a medium size acidic amino acid was substituted by alanine, a small size hydrophobic amino acid in *Bos Taurus*. On the 8th substitution, while *Bos Taurus*, *Equus caballus*, and *Canis familiaris* have a serine, it is substituted for asparagine in *Pan troglodytes*, and *Homo sapiens*, and for arginine in *Macaca mulatta* and *Rattus norvegicus*. Additionally *Mus musculus* has a histidine, and *Monodelphis domestica* has a lysine at this position. On the 10th substitution, while *Bos Taurus*, *Equus caballus*, *Canis familiaris*, and *Monodelphis domestica* have an alanine, a small size hydrophobic amino acid, *Pan troglodytes*, *Homo sapiens*, and *Macaca mulatta* have aspartate, a medium size acidic amino acid. Both *Rattus norvegicus* and *Mus musculus* have phenylalanine at this position. For the second HARP domain, there were 34 positions with amino acid substitutions in at least one of the species studied. These substitutions are highlighted in the alignment (Figure 6B). Again *Monodelphis domestica* was the most distant species for this domain.

In phylogenetic tree analysis of HARP1 domain, significantly higher bootstrap values were observed for *Rattus norvegicus*, *Mus musculus*, *Pan troglodytes* and *Homo sapiens*. In the second HARP domain, high bootstrap values conserved only in *Rattus norvegicus* and *Mus musculus*. For both domains *Monodelphis domestica* observed as the most distant mammalian among 9 species. When we compared first and second domain of HARP in SMARCAL1 also there was a separation which can easily be identified between the group of *Canis familiaris*, *Rattus norvegicus*, *Mus musculus* and the group of *Pan troglodytes*, *Homo sapiens*, and *Macaca mulatta*. *Equus caballus* was observed closer to the second group in the first HARP domain.

The phylogenetic tree of SNF2N which is the third domain of SMARCAL1 shows similar composition like the first two phylogenetic trees (Figure 7). But the most significant difference is *Canis familiaris* is getting closer to *Bos taurus*. For the last domain of SMARCAL1, one of the clearest observations is lowering of bootstrap values between *Bos taurus* and *GLEAN 20241*, when it is compared to the

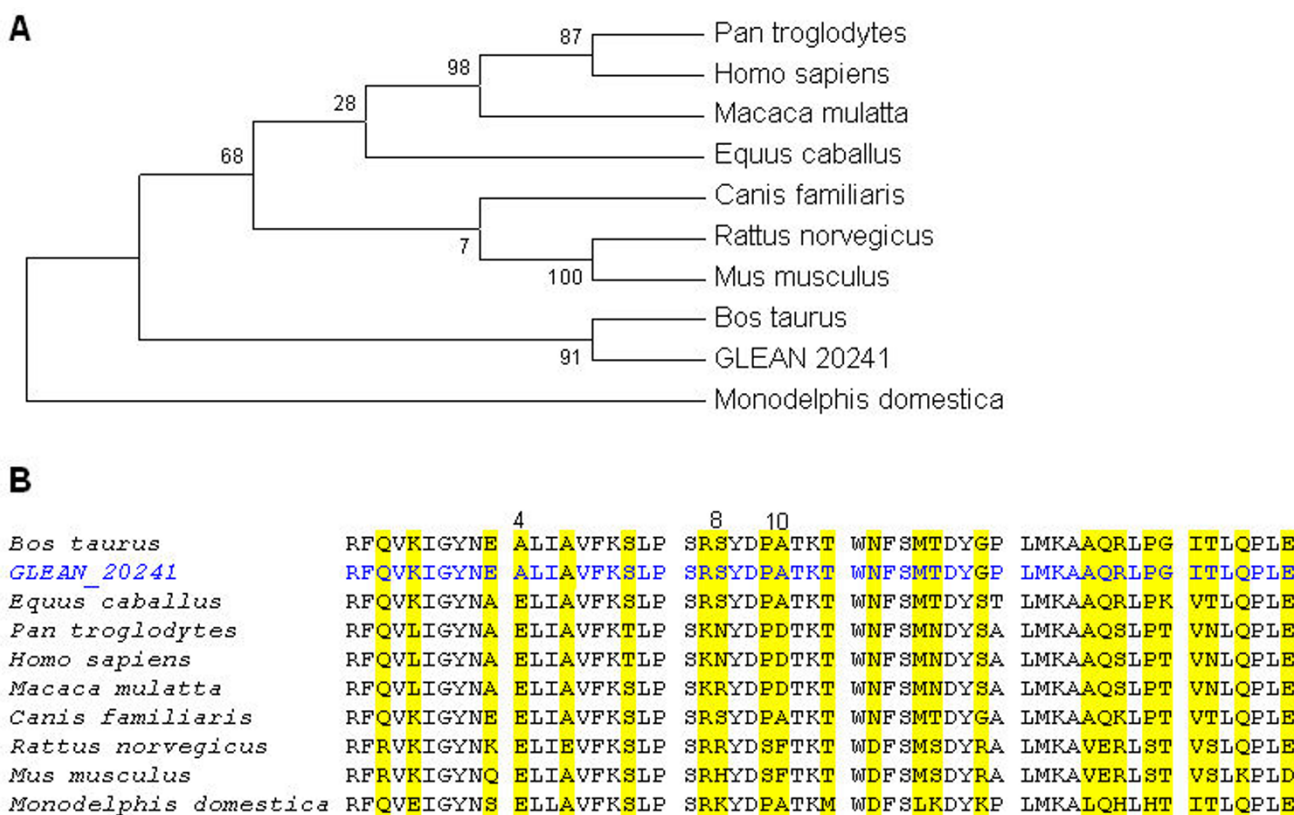


Figure 5
A. SMARCAL1 first HARP domain phylogenetic tree with the highest parsimony (length = 46). The consistency index is 0.95, the retention index is 0.94, and the composite index is 0.90 for all sites and parsimony-informative sites. There were a total of 57 positions in the final dataset, of which 20 were parsimony informative. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) is shown next to the branches. **B.** Multiple sequence alignment of the first HARP domain in SMARCAL1. Substitutions in at least one species are highlighted. Numbers on top of the alignment show significant substitutions which are mentioned in the text.

other phylogenetic trees. Also in the phylogenetic and sequence related analysis 11 parsimony informative sites detected and 46 of the sites are conserved among the species (Figure 8).

The multiple sequence alignment for the SNF2 family N-terminal domain is shown in Figure 8, with *Monodelphis domestica* as the most distant organism among the other mammals. Positions with insertions and deletions are marked with stars. The first insertion comprises 3 additional amino acids (glutamate, leucine, and lysine) present only in the *Equus caballus*, protein. There is a deletion of the amino acid arginine, present in all species, except for the NCBI bovine sequence. However the GLEAN_20241 does not have the deletion. The amino acid threonine is also absent in both *Rattus norvegicus* and *Mus musculus*. The bovine NCBI sequence showed significant mutations of the third conserved domain marked in

red on the alignment. However the sequenced official gene set for this protein (Bovine Genome Database <http://racerx00.tamu.edu/bovine>) shows a higher homology to all species, differing in only 2 amino acids from the horse and human protein. These findings indicate sequencing errors in the currently available bovine SMARCAL1 protein. These errors will likely be corrected with the completion of the bovine genome annotation effort. The bovine helicase C-terminal domain protein shows a deletion (marked with a star) and several substitutions highlighted in red (Figure 9B) that do not exist in GLEAN_20241. These observations point to the need for an update in SMARCAL1 protein sequence currently available at NCBI.

In addition to our analysis, we applied disparity index, ID [24], which measures the observed difference in evolutionary patterns for a pair of sequences. The disparity index for HMGN3a (Figure 10), did not show any signifi-

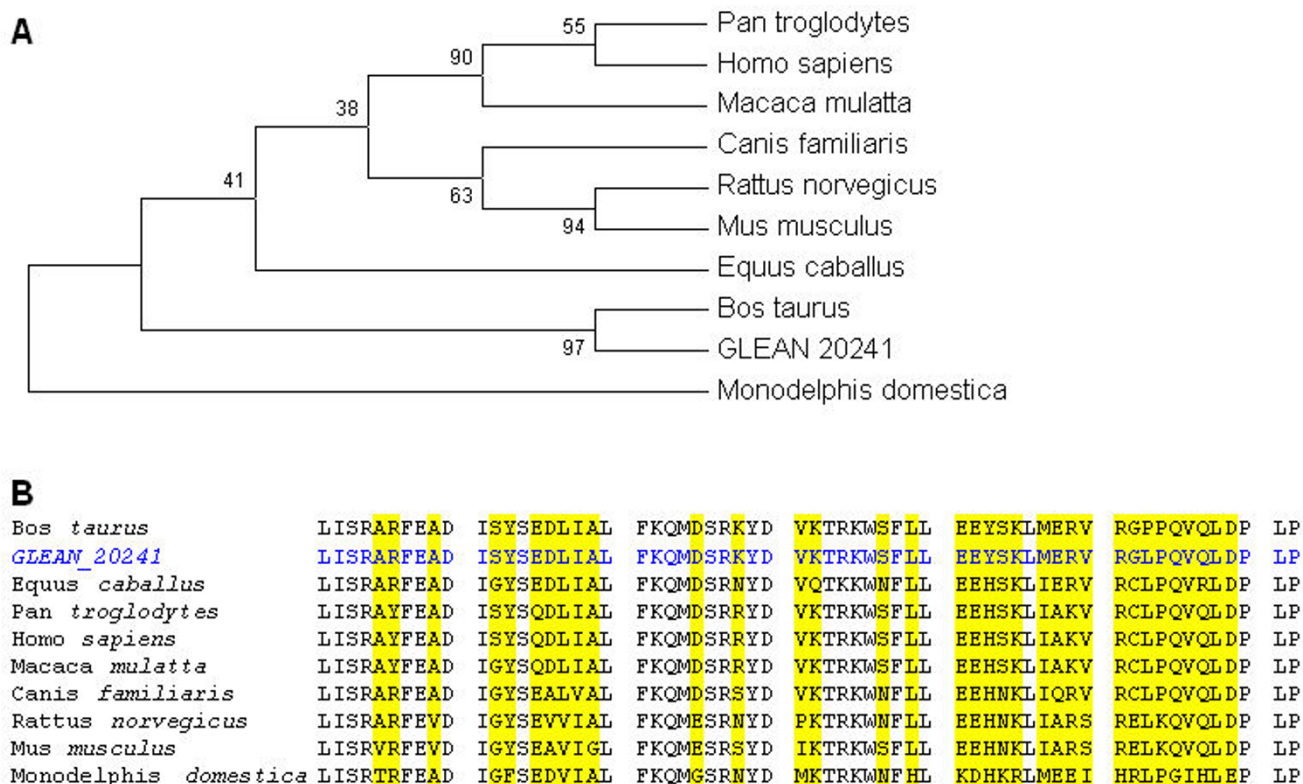


Figure 6
A. SMARCAL1 second HARP domain phylogenetic tree with the highest parsimony (length = 53). The consistency index is 0.84, the retention index is 0.85, and the composite index is 0.77 for all sites. There were a total of 62 positions in the final dataset, of which 17 were parsimony informative. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) is shown next to the branches. **B.** Multiple sequence alignment of the second HARP domain in SMARCAL1. Highlighted regions show substitutions occur.

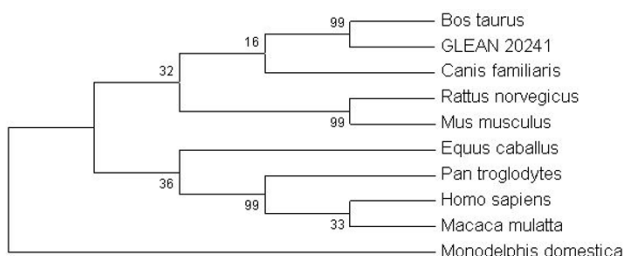


Figure 7
SMARCAL1 SNF2N domain phylogenetic tree with the highest parsimony (length = 145). The consistency index is 0.81, the retention index is 0.79, and the composite index is 0.70 for all sites. There were a total of 290 positions in the final dataset, out of which 39 were parsimony informative. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) is shown next to the branches.

cant pairs of species. The disparity index for each domain of SMARCAL1 is presented in Figure 11. In the first HARP domain (Figure 11A) 6 pairs of species (*Bos taurus*-*Rattus norvegicus*, *Pan troglodytes*-*Rattus norvegicus*, *Homo sapiens*-*Rattus norvegicus*, *Macaca mulatta*-*Rattus norvegicus*, *Canis familiaris*-*Rattus norvegicus*, and *Mus musculus*-*Homo sapiens*) were considered significant. The disparity index did not observed differences in evolutionary patterns for the second HARP domain (Figure 11B). There were 9 significant pairs in the SNF2 N-terminal domain disparity index (*Bos Taurus*-*Equus caballus*, *Bos Taurus*-*Pan troglodytes*, *Bos Taurus*-*Homo sapiens*, *Bos taurus* -*Macaca mulatta*, *Bos Taurus*-*Rattus norvegicus*, *Bos Taurus*-*Monodelphis domestica*, *Equus caballus*-*Pan troglodytes*, *Equus caballus*-*Homo sapiens*, *Equus caballus*-*Macaca mulatta*) (Figure (Figure 11C). In the disparity index for the helicase C-terminal domain only the pair *Bos taurus*-*Canis familiaris* was significant (Figure 11D).


```

Bos taurus          FQRAGVNFATAIQGRLLLADDMGLGKTIQAICTAAAYRKEWPELLVVVPS
GLEAN_20241        FQRAGVNFATAIQGRLLLADDMGLGKTIQAICTAAAYRKEWPELLVVVPS
Equus caballus     FQRAGVNFATAIQGRLLLADDMGLGKTIQAICTAAAYRKEWPELLVVVPS
Pan troglodytes   FQRAGVNFATAIQGRLLLADDMGLGKTIQAICTAAAYRKEWPELLVVVPS
Homo sapiens      FQRAGVNFATAIQGRLLLADDMGLGKTIQAICTAAAYRKEWPELLVVVPS
Macaca mulatta    FQRAGVNFATAIQGRLLLADDMGLGKTIQAICTAAAYRKEWPELLVVVPS
Canis familiaris  FQRAGVNFATAIQGRLLLADDMGLGKTIQAICTAAAYRKEWPELLVVVPS
Rattus norvegicus FQRAGVNFATAIQGRLLLADDMGLGKTIQAICTAAAYRKEWPELLVVVPS
Mus musculus      FQRAGVNFATAIQGRLLLADDMGLGKTIQAICTAAAYRKEWPELLVVVPS
Monodelphis domestica FQRIGVNFATAISKGRLLLADDMGLGKTIQAICTAAAYRTEWPELLVVVTPST

Bos taurus          VRFTWEQAFCRWLPSSLNPLDINVVVVTGKDRLLTDGLVNIVSFDLLSKLEKQ
GLEAN_20241        VRFTWEQAFCRWLPSSLNPLDINVVVVTGKDRLLTDGLVNIVSFDLLSKLEKQ
Equus caballus     VRFTWEQAFQWLPSSLNPLDINVVVVTGKDRLLTDGLVNIVSFDLLSKLEKQ
Pan troglodytes   VRFTWEQAFLRWLPSSLNPLDINVVVVTGKDRLLTDGLVNIVSFDLLSKLEKQ
Homo sapiens      VRFTWEQAFLRWLPSSLNPLDINVVVVTGKDRLLTDGLVNIVSFDLLSKLEKQ
Macaca mulatta    VRFTWEQAFLRWLPSSLNPLDINVVVVTGKDRLLTDGLVNIVSFDLLSKLEKQ
Canis familiaris  VRFTWEQAFLRWLPSSLNPLDINVVVVTGKDRLLTDGLVNIVSFDLLSKLEKQ
Rattus norvegicus VRFTWEQAFLRWLPSSLNPLDINVVVVTGKDRLLTDGLVNIVSFDLLSKLEKQ
Mus musculus      VRFTWEQAFLRWLPSSLNPLDINVVVVTGKDRLLTDGLVNIVSFDLLSKLEKQ
Monodelphis domestica VRFTWEQAFHRWLPSSLNPLDINVVVVTGKDRLLTDGLVNIVSFDLLSKMNNK

      ****
Bos taurus          LK---PPFKVVIIDESHFLKNIKTARCAAMPVLLKVAKRVIILSGTPAMS
GLEAN_20241        LK---PPFKVVIIDESHFLKNIKTARCAAMPVLLKVAKRVIILSGTPAMS
Equus caballus     LKQKLSPPFKVVIIDESHFLKNIKTARCAAMPVLLKVAKRVIILSGTPAMS
Pan troglodytes   LK---TPFKVVIIDESHFLKNIKTARCAAMPVLLKVAKRVIILSGTPAMS
Homo sapiens      LK---TPFKVVIIDESHFLKNIKTARCAAMPVLLKVAKRVIILSGTPAMS
Macaca mulatta    LK---TPFKVVIIDESHFLKNIKTARCAAMPVLLKVAKRVIILSGTPAMS
Canis familiaris  LK---TPFKVVIIDESHFLKNIKTARCAAMPVLLKVAKRVIILSGTPAMS
Rattus norvegicus LK---TPFKVVIIDESHFLKNIKTARCAAMPVLLKVAKRVIILSGTPAMS
Mus musculus      LK---TPFKVVIIDESHFLKNIKTARCAAMPVLLKVAKRVIILSGTPAMS
Monodelphis domestica LQ---TSFQVVIIDESHFLKNIKTARCAAMPVLLKVAKRVIILSGTPAMS

Bos taurus          RPAELYTQILA VRPTFFPQFHAFGLRYCGAKRQPMGWMDYSGSSNLGELKL
GLEAN_20241        RPAELYTQILA VRPTFFPQFHAFGLRYCGAKRQPMGWMDYSGSSNLGELKL
Equus caballus     RPAELYTQILA VRPTFFPQFHAFGLRYCGAKRQPMGWMDYSGSSNLGELKL
Pan troglodytes   RPAELYTQILA VRPTFFPQFHAFGLRYCGAKRQPMGWMDYSGSSNLGELKL
Homo sapiens      RPAELYTQILA VRPTFFPQFHAFGLRYCGAKRQPMGWMDYSGSSNLGELKL
Macaca mulatta    RPAELYTQILA VRPTFFPQFHAFGLRYCGAKRQPMGWMDYSGSSNLGELKL
Canis familiaris  RPAELYTQILA VRPTFFPQFHAFGLRYCGAKRQPMGWMDYSGSSNLGELKL
Rattus norvegicus RPAELYTQILA VRPTFFPQFHAFGLRYCGAKRQPMGWMDYSGSSNLGELKL
Mus musculus      RPAELYTQILA VRPTFFPQFHAFGLRYCGAKRQPMGWMDYSGSSNLGELKL
Monodelphis domestica RPAELYTQILA VRPTFFPQFHAFGLRYCGAKRQPMGWMDYSGSSNLGELKL

Bos taurus          LLEEAVMLRRLKSDVLSQLPAKQRKRMVVVAPGRINARTRAAALDAAAKEMT
GLEAN_20241        LLEEAVMLRRLKSDVLSQLPAKQRKRMVVVAPGRINARTRAAALDAAAKEMT
Equus caballus     LLEEAVMLRRLKSDVLSQLPAKQRKRMVVVAPGRINARTRAAALDAAAKEMT
Pan troglodytes   LLEEAVMLRRLKSDVLSQLPAKQRKRMVVVAPGRINARTRAAALDAAAKEMT
Homo sapiens      LLEEAVMLRRLKSDVLSQLPAKQRKRMVVVAPGRINARTRAAALDAAAKEMT
Macaca mulatta    LLEEAVMLRRLKSDVLSQLPAKQRKRMVVVAPGRINARTRAAALDAAAKEMT
Canis familiaris  LLEEAVMLRRLKSDVLSQLPAKQRKRMVVVAPGRINARTRAAALDAAAKEMT
Rattus norvegicus LLEEAVMLRRLKSDVLSQLPAKQRKRMVVVAPGRINARTRAAALDAAAKEMT
Mus musculus      LLEEAVMLRRLKSDVLSQLPAKQRKRMVVVAPGRINARTRAAALDAAAKEMT
Monodelphis domestica LLEEAVMLRRLKSDVLSQLPAKQRKRMVVVAPGRINARTRAAALDAAAKEMT

Bos taurus          TKDKTKQQQKEALILFFNRTAEAKIPSVIEYILDLLS GREKFLV
GLEAN_20241        TKDKTKQQQKEALILFFNRTAEAKIPSVIEYILDLLS GREKFLV
Equus caballus     TKDKTKQQQKEALILFFNRTAEAKIPCVIEYILDLLS GREKFLV
Pan troglodytes   TKDKTKQQQKDALILFFNRTAEAKIPSVIEYILDLLS GREKFLV
Homo sapiens      TKDKTKQQQKDALILFFNRTAEAKIPSVIEYILDLLS GREKFLV
Macaca mulatta    TMDKTKQQQKDALILFFNRTAEAKIPSVIEYILDLLS GREKFLV
Canis familiaris  TMDKTKQQQKEALILFFNRTAEAKIPSVIEYILDLLS GREKFLV
Rattus norvegicus -KDKTKQQQKEALLVFFNRTAEAKIPCVIEYILDLLS GREKFLV
Mus musculus      -KDKTKQQQKEALLVFFNRTAEAKIPCVIEYILDLLS GREKFLV
Monodelphis domestica TEKTSRRQQQKEALLVFFNRTAEAKIQPATDYILDLLS GREKFLV
    
```

Figure 8
Multiple sequence alignment of the SNF2N domain in SMARCAL1. Highlighted regions show where the substitutions occur. The multiple substitutions marked in red in the *Bos taurus* sequence may be due to sequencing errors since the corrected model for this protein (GLEAN_20241, in blue) added to this portion of the alignment only differs from the human sequence in 2 amino acids.

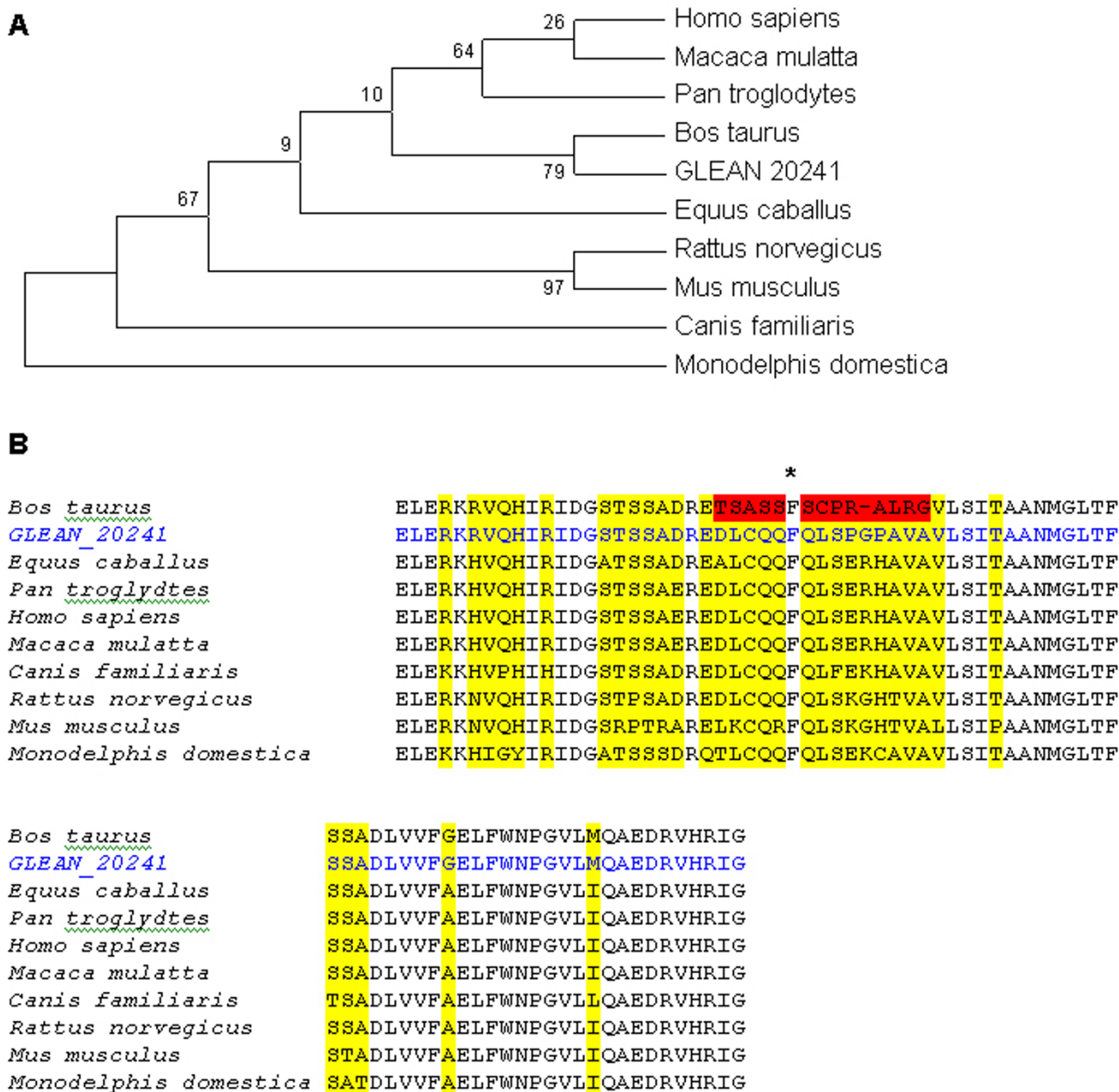


Figure 9

A. SMARCAL1 helicase C-terminal domain in phylogenetic tree with the highest parsimony (length = 52). The consistency index is 0.86, the retention index is 0.81, and the composite index is 0.77 for all sites. There were a total of 78 positions in the final dataset, out of which 11 were parsimony informative. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) is shown next to the branches. **B.** Multiple sequence alignment of the helicase C-terminal domain in SMARCAL1. Highlighted regions show where the substitutions occur.

Annotation of HMGN3a and SMARCAL1

The official gene model for HMGN3a (GLEAN_08006) was exactly identical to the NCBI protein (NP_001029676.1) with 6 exons and a total of 100 amino acids. No changes were annotated for this protein.

The official gene model for SMARCAL1 (GLEAN_20241) with 937 amino acids differed from the bovine NCBI protein (NP_788839.1) with 941 amino acids, particularly in 2 of the 16 exons. The official gene model for SMARCAL1 showed a higher level of sequence similarity to the human

	1	2	3	4	5	6	7	8
Bos taurus		0.000	0.000	0.000	0.021	0.000	0.000	0.011
GLEAN 08006	1.000		0.000	0.000	0.021	0.000	0.000	0.011
Equus caballus	1.000	1.000		0.000	0.000	0.000	0.000	0.000
Homo sapiens	1.000	1.000	1.000		0.000	0.000	0.000	0.000
Macaca ulata	0.318	0.333	1.000	1.000		0.000	0.000	0.000
Canis familiaris	1.000	1.000	1.000	1.000	1.000		0.011	0.000
Rattus norvegicus	1.000	1.000	1.000	1.000	1.000	0.402		0.074
Mus musculus	0.388	0.407	1.000	1.000	1.000	1.000	0.125	

Figure 10
Probability of rejecting the null hypothesis that HMG3a sequences have evolved with the same pattern of substitution, as judged from the extent of differences in base composition biases between sequences (Disparity Index test). A Monte Carlo test (1000 replicates) was used to estimate the *P*-values, which are shown below the diagonal. *P*-values smaller than 0.05 are considered significant. The estimates of the disparity index per site are shown for each sequence pair above the diagonal. There were a total of 95 positions in the final dataset. None of the *P*-values were smaller than 0.05. *All positions containing gaps and missing data were eliminated from the dataset.* **Black colored numbers:** Probability computed (must be < 0.05 for hypothesis rejection at 5% level), **Blue colored numbers:** Disparity Index.

SMARCAL1 protein (NP_054859.2) compared to the available bovine SMARCAL1 protein at NCBI. These results are presented in Table 1. The differences between GLEAN_20241 and NP_788839.1 were caused by an incorrect translation start and several sequencing errors that included 3 amino acid substitutions in exon three, one amino acid substitution in exon five, a sequencing error of 18 amino acids in exon ten, and a sequencing error of 13 amino acids in exon thirteen.

Annotation of SMARCAL1 mRNA showed that the first HARP conserved domain in SMARCAL1 is composed of the last part of the first exon, the second exon and the first part of the third exon. The second HARP conserved domain is composed of the end of the third exon, the fourth exon and the first part of the fifth exon. The third conserved domain, a helicase like domain named SNF2 family N-terminal domain is composed of the end of the fifth exon and exons 6–12. The fourth conserved domain, the Helicase C-terminal domain is composed of the last part of the twelfth exon and exons 13 and 14. The NCBI protein sequence for the Helicase C-terminal domain may also contain several sequencing errors.

Analysis of the conserved/non-conserved regions on the comparative modeled structure of SMARCAL1

Since the protein structures for SMARCAL1 are available for helicase like and helicase domain, a comparative homology model was built on covering only these domains. The percentage similarity between template and protein sequence was 24%. Depending on the multiple

sequence alignments, all non-conserved residues were mapped on the modeled structure (Figure 12). SNF2N and helicase C domains have nucleotide binding and ATP binding residues. There are 9 residues responsible for nucleotide binding and 8 residues for ATP binding which were retrieved from the literature [25,26]. These locations were mapped on the modeled structure. In our analysis we showed that all ATP binding residues exist in the conserved regions. Although Threonine781 is among the residues that are responsible in nucleotide binding, falls into the non-conserved region in protein sequence. Multiple alignment results show that in this specific location only one species (in *Mus musculus*) has variation which is Proline (Figure 12). This substitution creates a difference in amino acid side chain polarity as well as hydrophobicity and size at that specific position.

Conclusion

In our analysis, the bovine HMG3a and SMARCAL1 showed a higher degree of homology in all studied mammals. This high structural conservation highlights the importance of chromatin remodeling in the regulation of gene expression, particularly during early embryonic development. Understanding the interactions between these proteins and their roles could improve our understanding of epigenetics in reproduction and disease. Appropriate models for the study of chromatin remodeling proteins are essential to understanding this process, particularly in the case of diseases like Schimke immunosseous dysplasia (SIOD), caused by a mutation in the SMARCAL1 gene. The greater similarities of the HMG3a and SMARCAL1 proteins in human and bovine could suggest that more attention should be paid to a bovine model in the study of chromatin remodeling.

Methods

Production of bovine matured oocytes and early embryos in vitro

Chemicals and culture media

All chemicals were purchased from Sigma Chemical Company (St. Louis, MO, USA) unless otherwise stated. The synthetic oviduct fluid (SOF, Specialty Media) was used as a base media for embryo culture.

In vitro maturation (IVM)

Oocytes were collected from 2–8 mm follicles of bovine ovaries obtained from a local slaughterhouse in Wisconsin. Only oocytes containing several layers of cumulus cells and homogenous cytoplasm were selected. Oocytes were washed three times in TL-HEPES and matured in Tissue Culture Medium (TCM) 199 (Gibco/Invitrogen) supplemented with 0.2 mM pyruvate, 0.5 µg/ml follicle-stimulating hormone (FSH; Sioux Biochemicals, Sioux City, IA, USA), 5 µg/ml luteinizing hormone (LH; Sioux Biochemicals), 10% fetal calf serum (FCS, Gibco/Invitro-

	1	2	3	4	5	6	7	8	9	10
A										
<i>Bos taurus</i>		0.000	0.000	0.000	0.000	0.000	0.000	0.211	0.105	0.035
GLEAN 20241	1.000		0.000	0.000	0.000	0.000	0.000	0.211	0.105	0.035
<i>Equus caballus</i>	1.000	1.000		0.000	0.000	0.000	0.000	0.105	0.000	0.000
<i>Pan troglodytes</i>	1.000	1.000	1.000		0.000	0.000	0.000	0.368	0.211	0.000
<i>Homo sapiens</i>	1.000	1.000	1.000	1.000		0.000	0.000	0.368	0.211	0.000
<i>Macaca mulatta</i>	1.000	1.000	1.000	1.000	1.000		0.000	0.211	0.053	0.000
<i>Canis familiaris</i>	1.000	1.000	1.000	1.000	1.000	1.000		0.211	0.123	0.000
<i>Rattus norvegicus</i>	0.039	0.041	0.145	0.002	0.002	0.028	0.021		0.000	0.140
<i>Mus musculus</i>	0.198	0.162	1.000	0.056	0.046	0.339	0.106	1.000		0.035
<i>Monodelphis domestica</i>	0.354	0.334	1.000	1.000	1.000	1.000	1.000	0.142	0.372	
B										
<i>Bos taurus</i>		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.081
GLEAN 20241	1.000		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.113
<i>Equus caballus</i>	1.000	1.000		0.000	0.000	0.000	0.000	0.000	0.000	0.000
<i>Pan troglodytes</i>	1.000	1.000	1.000		0.000	0.000	0.000	0.032	0.000	0.161
<i>Homo sapiens</i>	1.000	1.000	1.000	1.000		0.000	0.000	0.032	0.000	0.161
<i>Macaca mulatta</i>	1.000	1.000	1.000	1.000	1.000		0.000	0.016	0.000	0.097
<i>Canis familiaris</i>	1.000	1.000	1.000	1.000	1.000	1.000		0.000	0.000	0.065
<i>Rattus norvegicus</i>	1.000	1.000	1.000	0.365	0.349	0.406	1.000		0.000	0.000
<i>Mus musculus</i>	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		0.000
<i>Monodelphis domestica</i>	0.263	0.197	1.000	0.153	0.136	0.242	0.289	1.000	1.000	
C										
<i>Bos taurus</i>		0.110	0.166	0.138	0.138	0.128	0.052	0.121	0.100	0.193
GLEAN 20241	0.006		0.038	0.000	0.000	0.000	0.000	0.014	0.000	0.038
<i>Equus caballus</i>	0.010	0.145		0.072	0.076	0.059	0.003	0.000	0.000	0.007
<i>Pan troglodytes</i>	0.021	1.000	0.011		0.000	0.000	0.000	0.000	0.000	0.000
<i>Homo sapiens</i>	0.016	1.000	0.009	1.000		0.000	0.000	0.000	0.000	0.000
<i>Macaca mulatta</i>	0.029	1.000	0.016	1.000	1.000		0.000	0.003	0.000	0.000
<i>Canis familiaris</i>	0.165	1.000	0.078	1.000	1.000	1.000		0.000	0.000	0.007
<i>Rattus norvegicus</i>	0.035	0.318	0.441	1.000	1.000	0.400	1.000		0.000	0.034
<i>Mus musculus</i>	0.074	1.000	1.000	1.000	1.000	1.000	1.000	1.000		0.041
<i>Monodelphis domestica</i>	0.039	0.259	0.389	1.000	1.000	1.000	0.422	0.280	0.291	
D										
<i>Bos taurus</i>		0.013	0.128	0.051	0.051	0.051	0.244	0.051	0.026	0.051
GLEAN 20241	0.383		0.064	0.038	0.038	0.038	0.103	0.000	0.000	0.000
<i>Equus caballus</i>	0.064	0.076		0.000	0.000	0.000	0.051	0.026	0.013	0.026
<i>Pan troglodytes</i>	0.270	0.135	1.000		0.000	0.000	0.064	0.013	0.026	0.000
<i>Homo sapiens</i>	0.259	0.125	1.000	1.000		0.000	0.064	0.013	0.026	0.000
<i>Macaca mulatta</i>	0.243	0.125	1.000	1.000	1.000		0.064	0.013	0.026	0.000
<i>Canis familiaris</i>	0.013	0.031	0.134	0.068	0.059	0.061		0.000	0.000	0.000
<i>Rattus norvegicus</i>	0.254	1.000	0.244	0.346	0.371	0.359	1.000		0.013	0.000
<i>Mus musculus</i>	0.382	1.000	0.433	0.325	0.340	0.349	1.000	0.380		0.000
<i>Monodelphis domestica</i>	0.281	1.000	0.318	1.000	1.000	1.000	1.000	1.000	1.000	

Figure 11
Probability of rejecting the null hypothesis that the sequences of the SMARCAL1 conserved domains have evolved with the same pattern of substitution, as judged from the extent of differences in base composition biases between sequences (Disparity Index test). A Monte Carlo test (1000 replicates) was used to estimate the P-values, which are shown below the diagonal. P-values smaller than 0.05 are considered significant. The estimates of the disparity index per site are shown for each sequence pair above the diagonal. All positions containing gaps and missing data were eliminated from the dataset. **A.** First HARP domain: there were a total of 57 positions in the final dataset. **B.** Second HARP domain: there were a total of 62 positions in the final dataset. **C.** SNF2N domain: there were a total of 290 positions in the final dataset. **D.** Helicase C-terminal domain: there were a total of 78 positions in the final dataset. **Black colored numbers:** Probability computed (must be < 0.05 for hypothesis rejection at 5% level [yellow background]), **Blue colored numbers:** Disparity Index.

gen), 100 U/ml penicillin and 100 mg/ml streptomycin (Gibco/Invitrogen). Ten oocytes in each 50 µl maturation drop were covered with mineral oil and incubated for 24 h at 39 °C in a humidified incubator with 5% CO₂ [27]. After 24 hours, mature oocytes were washed twice with TL-HEPES. Mature oocytes were randomly selected for either RNA isolation or fertilization. Pools of 100 oocytes were frozen at -80 °C on RLT lysis buffer (Qiagen Valencia, CA) until RNA isolation.

In vitro fertilization (IVF)

Groups of 10 oocytes washed with TL-HEPES were transferred into 44 µl drops of fertilization medium (glucose-free TALP supplemented with 0.2 mM pyruvate, 6 mg/ml fatty acid-free BSA, 100 U/ml penicillin and 100 mg/ml streptomycin). Percoll gradient was used for separation of live spermatozoa in frozen-thawed semen [28]. Briefly, sperm was thawed at 36 °C for 1 min, and then carefully layered on top of the Percoll gradient system. Sperm was diluted in TL-HEPES to 5.0 × 10⁷ cells/ml and 2 µl of diluted sperm were added to the 44 µl fertilization drops, which produced a final sperm concentration of 2.0 × 10⁶ cell/ml. Fertilization drops (50 µl) were supplemented with 2 µl of 5 µg/ml heparin and 2 µl of PHE solution (20 µM penicillamine, 10 µM hypotaurine, 1 µM epinephrine) and 2 µl of semen (50 × 10⁶ sperm cells/ml) into the 44 µl fertilization drops [29].

In Vitro Culture (IVC)

Following 18 hours co-culture of oocytes and sperm, cumulus cells were removed by vortexing the presumptive zygotes in a 1.5 ml Eppendorf tube at high speed for 3 minutes. The cumulus free presumptive zygotes were washed three times in TL-HEPES and approximately 30 cumulus free presumptive zygotes transferred into a 50 µl drops (SOF) under mineral oil for embryo culture. At 72 hpi, 10% FCS was supplemented to each drop except that the 2–4-cell stage embryos were collected earlier than serum addition. In this study, the amount of embryos cleaved was 76.0% while the amount developing to the blastocyst stages was 21.8%.

Collection of embryos

Developing embryos of 2–4-cell, 8–16-cell, morulae and blastocysts stages were collected at 44, 100, 120, and 168

hpi, respectively. At the beginning of the embryo culture, the drops were randomly assigned to each developmental stages to collect embryos at the corresponding time. Therefore, the embryos were collected from each drop only for one developmental stage mentioned above. Once the embryos were removed for a specific cell stage, the drops were crossed out to prevent duplicate collection from the same drop. Embryos developing to the corresponding stage were removed from culture drops and washed four times with TL-Hepes with PVP (3 mg/ml). The number of embryos pooled for the development stages of 2–4-cell, 8–16-cell, morulae and blastocysts stages were 50, 50, 20 and 5 per tube, respectively. Total of four tubes were collected from 2 replicates for each development stages. The pooled embryos were frozen at -80 °C on RLT lysis buffer (Qiagen Valencia, CA) until RNA isolation.

Isolation of RNA

Total RNA was isolated from pools of 100 oocytes, 100 2- to 4-cell embryos, 100 8- to 16-cell embryos and 10 expanded blastocysts (evaluated and graded according the International Embryo Transfer Society (IETS) guidelines [30]. Total RNA was isolated using an RNeasy Micro Kit (Qiagen) according to the manufacturer's instructions. Quality of total RNA was estimated using the Bioanalyzer 2100 RNA 6000 picochip kit (Agilent, Palo Alto, CA, USA). RNA quantity and purity were determined using a NanoDrop® ND-1000 Spectrophotometer (NanoDrop Technologies, Wilmington, DE). Total RNA from all groups was normalized to 4 ng and used for cDNA synthesis using SuperScript III Platinum Two Step qRT-PCR kit according to the manufacturer's protocol. Cycling temperatures and times were 25 °C for 10 min, 42 °C for 50 min, and 85 °C for 5 min.

Real time reverse transcriptase PCR to determine transcript abundance of HMGN3a and SMARCAL1

Complementary DNA was generated using the SuperScript III Platinum® Two-Step qRT-PCR Kit (Invitrogen Life Technologies, Carlsbad, CA) according to the manufacturer's protocol. The kit includes both oligo(dT)₂₀ and random hexamers for cDNA generation. Primers were designed using Primer Premier 5 software (Premier Bio-soft International, Palo Alto, CA, USA) spanning 2 exons

Table 1: Pairwise alignment results comparing both the NCBI bovine SMARCAL1 protein and the official gene model for SMARCAL1 (GLEAN_20241) to the human SMARCAL1 protein.

	Bovine NCBI SMARCAL1 vs. Human SMARCAL1	GLEAN_20241 vs. Human SMARCAL1
Similarity Score	3483	3799
Match	74%	79%
Number of Matches	718	764
Number of Mismatches	208	165
Total Length of Gaps	43	33

The official gene model shows a higher sequence homology to the human protein with more matches, shorter gaps, and fewer mismatches.

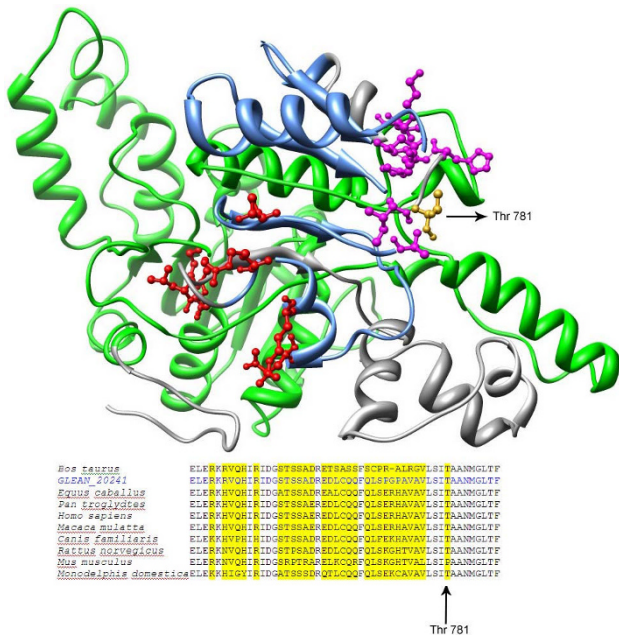


Figure 12
Based on the domains in the multiple sequence alignment, residues are colored in green and blue. Green color represents the helicase like domain and blue color represents helicase domain. ATP binding residues are shown in red color and represented as balls and sticks. Nucleotide binding residues are colored in magenta color. The arrows shows the residue which existed in the non-conservative region on the structure and in the MSA, which is indicated in yellow.

to avoid genomic DNA amplification (Table 2). Real-time quantitative PCR was performed to assess transcript abundance of HMGN3a and SMARCAL1 relative to the housekeeping gene GAPDH. Quantitative assessment of RNA amplification was detected by SYBR® GreenER™ qPCR SuperMixes for iCycler (Invitrogen Life Technologies, Carlsbad, CA, 11761-100). Five µl cDNA were used for quantitative Real-time PCR reactions according to the iCycler iQ Real-Time PCR instrument (BIO-RAD). Primer concentration was adjusted to 10 µM. The cycling parameters were 50 °C for 2 min, 95 °C for 8 min 30 sec for denaturation, 40 cycles of 15 sec at 95 °C, and 30 sec at 60 °C

and 30 sec at 72 °C for amplification and extension. The melting curve was performed starting at 55 °C with 0.5 °C increase for 10 sec in 80 cycles. Expression values were calculated using the relative standard curve method. Standard curves were generated using 10-fold serial dilutions for GAPDH and the target gene by measuring the cycle number at which exponential amplification occurred. Results from different groups were analyzed by one-way analysis of variance (ANOVA) by SAS 9.1.3 (SAS Institute inc. Carey, NC). Relative expression software tool (REST) was used to compare all samples of each group. The mathematical model used by REST is based on the PCR efficiencies and the crossing point deviation between samples [31].

Comparative functional genomics analyses of HMGN3a and SMARCAL1 across mammals

We retrieved the protein sequences of SMARCAL1 from NCBI by performing protein BLAST [32] against mammalian database using *Bos taurus* SMARCAL1 (NP_788839) as our query protein. Sequence data were manipulated with the Friend software [33], a bioinformatics application designed for simultaneous analysis and visualization of multiple structures and sequences of proteins, DNA or RNA. Multiple sequence alignment of 9 mammalian SMARCAL1 protein sequences that are listed in Table 3 were created by using Clustal W [34] under Friend Software. We defined conserved regions based on domains listed in the Pfam [35] database which has conserved amino acid sequence regions. The same steps were applied for constructing HMGN3a phylogenetic tree, for which we used the only available Reference Sequence protein for *Bos Taurus* (NP_001029676). Since the availability of mammalian HMGN3a sequences is limited, we excluded *Monodelphis domestica* from the HMGN3a phylogenetic analyses, which were conducted in MEGA 4 [24]. There are 7 mammalian species used in our analysis of HMGN3a which are shown in Table 4. The Maximum Parsimony method was used for inferring the evolutionary history when creating the phylogenetic trees for both SMARCAL1 and HMGN3a.

Comparative modeling of SMARCAL1

Since the availability of possible templates for comparative modeling, model was created for only SMARCAL1.

Table 2: Primers used for Real Time PCR gene expression analysis of HMGN3a, the housekeeping gene GAPDH, and SMARCAL1.

Gene	Primer sequence and position (5' - 3')	Product size (bp)	Accession Number
HMGN3a_F	GTTCCAGCCC GTT GCTTTAC (22 – 42)	355	NM_001034504.1
HMGN3a_R	GACCATTCATTCTCCCTCGTTAC (376 – 399)		
GAPDH_F	TGCTGGTGCTGAGTATGTGGT (333 – 354)	295	XM_865742
GAPDH_R	AGTCTTCTGGGTGGCAGTGAT (627–648)		
SMARCAL1_F	CCATCTGCATAGCGGCCTAT (1397–1416)	141	NM_176666
SMARCAL1_R	CGGTTACCACGACGTTGATGT (1537–1517)		

Table 3: Organisms and protein accession numbers used in multiple sequence alignment of SMARCAL1.

Organism	Protein accession id
<i>Bos taurus</i>	NP_788839
<i>Homo sapiens</i>	NP_054859
<i>Equus caballus</i>	XP_001490055
<i>Pan troglodytes</i>	XP_516076
<i>Macaca mulatta</i>	XP_001086469
<i>Canis familiaris</i>	XP_536062
<i>Rattus norvegicus</i>	NP_001101692
<i>Mus musculus</i>	NP_061287
<i>Monodelphis domestica</i>	XP_001367889

PDB-file 1z63 chain A was used as a template for comparative modeling which was sharing 24% with SMARCAL1 protein sequence of *Bos taurus*. The template that was including the residues from 422 to 869 covered two domains of SMARCAL1 and these were helicase like and helicase domains. Comparative modeling was performed by Modeller 9v1 [36]. Structural analysis was done under Friend and model picture was created with Chimera [37].

Annotation of HMG3a and SMARCAL1

Annotations were performed using the Apollo software, an interactive tool that enables gene annotators to inspect computationally obtained gene predictions, and edit them by evaluating all the data supporting each annotation [38]. Apollo was successfully used to annotate the *Drosophila melanogaster* genome [39], and was the tool recommended by the Bovine Genome Sequencing Consortium for manual annotation of bovine genes. Apollo software was used to confirm the protein sequence accuracy by transcribing, and translating the DNA, identifying untranslated regions (UTR), translation start, exons, introns, and translation stops. Previous protein information from NCBI or Ensembl, was compared to the GLEAN sequence and errors in the proteins were analyzed in detail. The GLEAN identification number for HMG3a was GLEAN_08006, and for SMARCAL1 was GLEAN_20241. Annotation of SMARCAL1 and HMG3a were submitted to the Bovine Genome Annotation Submission Database at BovineGenome.org.

Table 4: Organisms and protein accession numbers used in multiple sequence alignment of HMG3a.

Organism	Protein accession id
<i>Bos taurus</i>	NP_001029676
<i>Homo sapiens</i>	NP_004233
<i>Equus caballus</i>	XP_001499146.2
<i>Macaca mulatta</i>	XP_001110668
<i>Canis familiaris</i>	XP_854051
<i>Rattus norvegicus</i>	NP_001007021
<i>Mus musculus</i>	NP_080398

Authors' contributions

AU performed comparative functional genomics analyses, and drafted the manuscript. NRO designed primers for HMG3a and SMARCAL1, annotated HMG3a and SMARCAL1, and drafted the manuscript. HW performed RNA isolations, Real Time RT-PCR experiments and analysis of the results. AK performed in vitro fertilization and culture of bovine oocytes and embryos, and collected the embryos for RNA isolation, and involved with drafting the manuscript. JJP participated in the design of the study, and in the *in vitro* production of the embryos, and reviewed the manuscript. VI participated in the design of the study, in its coordination and helped to draft the manuscript. EM participated in the design of the study, in its coordination and helped to draft the manuscript.

Acknowledgements

This study was partially supported by Mississippi Agricultural and Forestry Experiment Station (J-11470) and by National Library of Medicine (Grant Number: R01LM009519 to VI).

References

- Memili E, Dominko T, First NL: **Onset of transcription in bovine oocytes and preimplantation embryos.** *Mol Reprod Dev* 1998, **51(1)**:36-41.
- Memili E, First NL: **Control of gene expression at the onset of bovine embryonic development.** *Biol Reprod* 1999, **61(5)**:1198-1207.
- Misirlioglu M, Page GP, Sagirkaya H, Kaya A, Parrish JJ, First NL, Memili E: **Dynamics of global transcriptome in bovine matured oocytes and preimplantation embryos.** *Proc Natl Acad Sci USA* 2006, **103(50)**:18905-18910.
- Thompson EM, Legouy E, Renard JP: **Mouse embryos do not wait for the MBT: chromatin and RNA polymerase remodeling in genome activation at the onset of development.** *Dev Genet* 1998, **22(1)**:31-42.
- Olave I, Wang W, Xue Y, Kuo A, Crabtree GR: **Identification of a polymorphic, neuron-specific chromatin remodeling complex.** *Genes Dev* 2002, **16(19)**:2509-2517.
- Shirakawa H, Herrera JE, Bustin M, Postnikov Y: **Targeting of high mobility group-14/-17 proteins in chromatin is independent of DNA sequence.** *The Journal of biological chemistry* 2000, **275(48)**:37937-37944.
- Bustin M, Reeves R: **High-mobility-group chromosomal proteins: architectural components that facilitate chromatin function.** *Prog Nucleic Acid Res Mol Biol* 1996, **54**:35-100.
- West KL, Ito Y, Birger Y, Postnikov Y, Shirakawa H, Bustin M: **HMG3a and HMG3b, two protein isoforms with a tissue-specific expression pattern, expand the cellular repertoire of nucleosome-binding proteins.** *The Journal of biological chemistry* 2001, **276(28)**:25959-25969.
- Mohamed OA, Bustin M, Clarke HJ: **High-mobility group proteins 14 and 17 maintain the timing of early embryonic development in the mouse.** *Dev Biol* 2001, **229(1)**:237-249.
- Crippa MP, Nickol JM, Bustin M: **Developmental changes in the expression of high mobility group chromosomal proteins.** *The Journal of biological chemistry* 1991, **266(5)**:2712-2714.
- Korner U, Bustin M, Scheer U, Hock R: **Developmental role of HMG proteins in *Xenopus laevis*.** *Mech Dev* 2003, **120(10)**:1177-1192.
- Banine F, Bartlett C, Gunawardena R, Muchardt C, Yaniv M, Knudsen ES, Weissman BE, Sherman LS: **SWI/SNF chromatin-remodeling factors induce changes in DNA methylation to promote transcriptional activation.** *Cancer Res* 2005, **65(9)**:3542-3547.
- Fry CJ, Peterson CL: **Chromatin remodeling enzymes: who's on first?** *Curr Biol* 2001, **11(5)**:R185-197.
- Pollard KJ, Peterson CL: **Chromatin remodeling: a marriage between two families?** *Bioessays* 1998, **20(9)**:771-780.

15. Lusser A, Kadonaga JT: **Chromatin remodeling by ATP-dependent molecular machines.** *Bioessays* 2003, **25(12)**:1192-1200.
16. Zheng P, Patel B, McMenamin M, Paprocki AM, Schramm RD, Nagl NG Jr, Wilsker D, Wang X, Moran E, Latham KE: **Expression of genes encoding chromatin regulatory factors in developing rhesus monkey oocytes and preimplantation stage embryos: possible roles in genome activation.** *Biol Reprod* 2004, **70(5)**:1419-1427.
17. LeGouy E, Thompson EM, Muchardt C, Renard JP: **Differential pre-implantation regulation of two mouse homologues of the yeast SWI2 protein.** *Dev Dyn* 1998, **212(1)**:38-48.
18. Magnani L, Cabot RA: **Developmental arrest induced in cleavage stage porcine embryos following microinjection of mRNA encoding Brahma (Smarca 2), a chromatin remodeling protein.** *Mol Reprod Dev* 2007, **74(10)**:1262-1267.
19. Gebuhr TC, Bultman SJ, Magnuson T: **Pc-G/trx-G and the SWI/SNF connection: developmental gene regulation through chromatin remodeling.** *Genesis* 2000, **26(3)**:189-197.
20. Reyes JC, Barra J, Muchardt C, Camus A, Babinet C, Yaniv M: **Altered control of cellular proliferation in the absence of mammalian brahma (SNF2alpha).** *EMBO J* 1998, **17(23)**:6979-6991.
21. Coleman MA, Eisen JA, Mohrenweiser HW: **Cloning and characterization of HARP/SMARCAL1: a prokaryotic HepA-related SNF2 helicase protein from human and mouse.** *Genomics* 2000, **65(3)**:274-282.
22. Boerkoel CF, Takashima H, John J, Yan J, Stankiewicz P, Rosenbarker L, Andre JL, Bogdanovic R, Burguet A, Cockfield S, et al.: **Mutant chromatin remodeling protein SMARCAL1 causes Schimke immuno-osseous dysplasia.** *Nat Genet* 2002, **30(2)**:215-220.
23. Sun F, Tang F, Yan AY, Fang HY, Sheng HZ: **Expression of SRG3, a chromatin-remodelling factor, in the mouse oocyte and early preimplantation embryos.** *Zygote* 2007, **15(2)**:129-138.
24. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.** *Mol Biol Evol* 2007, **24(8)**:1596-1599.
25. Theis K, Chen PJ, Skorvaga M, Van Houten B, Kisker C: **Crystal structure of UvrB, a DNA helicase adapted for nucleotide excision repair.** *The EMBO journal* 1999, **18(24)**:6899-6907.
26. Muthuswami R, Truman PA, Mesner LD, Hockensmith JW: **A eukaryotic SWI2/SNF2 domain, an exquisite detector of double-stranded to single-stranded DNA transition elements.** *The Journal of biological chemistry* 2000, **275(11)**:7648-7655.
27. Sagirkaya H, Misirlioglu M, Kaya A, First NL, Parrish JJ, Memili E: **Developmental and molecular correlates of bovine preimplantation embryos.** *Reproduction* 2006, **131(5)**:895-904.
28. Parrish JJ, Krogenaes A, Susko-Parrish JL: **Effect of bovine sperm separation by either swim-up or Percoll method on success of in vitro fertilization and early embryonic development.** *Theriogenology* 1995, **44(6)**:859-869.
29. Leibfried ML, Bavister BD: **Effects of epinephrine and hypotaurine on in-vitro fertilization in the golden hamster.** *J Reprod Fertil* 1982, **66(1)**:87-93.
30. Stringfellow DA, Seidel SM: **Manual of the International Embryo Transfer Society: a Procedural Guide and General Information for the Use of Embryo Transfer Technology Emphasizing Sanitary Procedures.** 3rd edition. Savoy, Illinois: International Embryo Transfer Society; 1998.
31. Pfaffl MW, Horgan GW, Dempfle L: **Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR.** *Nucleic Acids Res* 2002, **30(9)**:e36.
32. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215(3)**:403-410.
33. Abyzov A, Errami M, Leslin CM, Ilyin VA: **Friend, an integrated analytical front-end application for bioinformatics.** *Bioinformatics* 2005, **21(18)**:3677-3678.
34. Thompson JD, Higgins DG, Gibson TJ: **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(22)**:4673-4680.
35. Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, et al.: **Pfam: clans, web tools and services.** *Nucleic Acids Res* 2006:D247-251.
36. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A: **Comparative protein structure modeling of genes and genomes.** *Annual review of biophysics and biomolecular structure* 2000, **29**:291-325.
37. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE: **UCSF Chimera – a visualization system for exploratory research and analysis.** *Journal of computational chemistry* 2004, **25(13)**:1605-1612.
38. Lewis SE, Searle SM, Harris N, Gibson M, Lyer V, Richter J, Wiel C, Bayraktaroglu L, Birney E, Crosby MA, et al.: **Apollo: a sequence annotation editor.** *Genome Biol* 2002, **3(12)**:RESEARCH0082.
39. Drysdale R: **The Drosophila melanogaster genome sequencing and annotation projects: a status report.** *Brief Funct Genomic Proteomic* 2003, **2(2)**:128-134.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

