Sequence Characterization of Promoter Region at the Melanocortin-1 Receptor (MC1R) Gene in Karakul Sheep Breed

M. Muhaghegh-Dolatabady1,∗ and J. Habibizad1

ABSTRACT

The Melanocortin-1 Receptor MC1R is encoded by the extension locus, playing a fundamental role in the determination of coat color in a number of mammalian species. However, so far there has been no report regarding the Single Nucleotide Polymorphisms (SNPs) of the MC1R promoter region and the potential association of its SNPs with coat color in sheep (Ovis aries). Throughout the present study, the promoter region of the MC1R gene was screened using Single Strand Conformation Polymorphism SSCP and DNA sequencing in the Karakul breed of sheep. A total of 4 distinct SSCP patterns were observed which revealed 3 novel SNPs and an insertion/deletion of 26 nucleotides upon sequence analysis in the analyzed population. In silico analysis of the MC1R promoter sequence predicted no consensus TATA-box motif at an appropriate position but detected multiple putative transcription factor binding sites for Ets, AML-1a, NF-E2, MZF1, USF, Oct-1 and GATA-1. The analysis of identified polymorphic sites also showed that the polymorphism at nucleotide position -89 relative to the start codon abolishes the USF transcription factor binding site. The SNP identified at the -100 position is located within a putative AML-1a transcription factor binding site. The insertion of 26 nucleotides at position -126 made a putative binding site for the MOK2 transcription factor. The possible functional activity of the identified genetic variations could be confirmed using gene expression analysis.

Keywords: Karakul sheep breed, MC1R, Promoter region, SNPs, Transcription factors.

INTRODUCTION

Coat color in livestock is highly varied, and bears significant biological and economical impact (Chen et al., 2009). Coat color is a highly heritable characteristic, playing a vital role in productive adaptability of livestock species (Peters et al., 1982). Understanding the genetic determinants of coat color in such livestock species as sheep, is critical for improving efficient selection of the desired trait. In mammals, coat color is mainly determined by the relative quantity of two basic melamins, eumelanin (black/brown) and phaeomelanin (yellow/red), which are genetically controlled by the Extension and Agouti loci, respectively (Jackson, 1994). The wild alleles at the two loci would result in different coat colors, whereas mutations can alter the eumelanin/phaeomelanin ratio and consequently cause coat color variations (Jackson, 1997). The extension locus encodes the Melanocortin 1 Receptor MC1R, which is a seven-transmembrane G-protein coupled receptor expressed on the surface of melanocytes and melanoma cells (Abdel-Malek, 2001; Robbins et al., 1993). MC1R gene mutations are associated with different coat color phenotypes in such livestock species as cattle (Klungland et al.,

1 Department of Animal Science, Faculty of Agriculture, Yasouj University, Yasouj, Islamic Republic of Iran.
∗ Corresponding author; e-mail: mmuhaghegh@mail.yu.ac.ir

551
Dominant MC1R mutations induce enhancement of eumelanin production while recessive inactivation mutations of MC1R stimulate phaeomelanin synthesis (Robbins et al., 1993). In sheep, MC1R gene is located on chromosome 14, and classical genetic studies have identified two alleles at the Extension locus: the dominant black (E\textsuperscript{D}) allele that may account for the black in a few breeds; and the wild type (E\textsuperscript{+}) allele widely distributed in most breeds in which segregation of Agouti alleles accounts for the majority of color variations (Russell, 1968; Sponenberg, 1997). In laboratory mice, MC1R is epistatic to ASIP, for example, dominant mutations in MC1R that lead to a constitutively active receptor are not inhibited by ASIP (Ollmann et al., 1998).

Karakul sheep is probably one of the oldest breeds of domesticated sheep in the world (Scherf et al., 2006). Coat color is economically important in Karakul and fine-fleeced sheep breeding, since it determines the value of the pelt, fur or wool. The Karakul breed is important for the production of lamb pelts of various colours for the fashion industry (Lundie, 2011). The coat color in Karakul pelts is mainly black, grey, white and brown. In general, the percentage of black pelt has increased, with a decline in demand for other colours (Schoeman, 1998). Several loci contribute to the expression of these coat colors in Karakul sheep with the E locus being mainly responsible for the expression of black color (Adalsteinsson, 1980; 1983). Two missense mutations were identified in the MC1R gene of sheep that determine dominant black (E\textsuperscript{D}) (Vage et al., 1999). The existence of these two mutations was also confirmed in Karakul sheep (Vage et al., 2003).

The promoter region of a gene, through binding of a specific transcription factor, is directly involved in gene transcription initiation, therefore sequence variation in this region may alter transcription factor binding sites, which in turn can affect gene expression and exert biological impacts. To date, there have been no reports on sequence variation in the promoter region of the ovine MC1R gene. Therefore, the objective followed in the present study was to identify sequence variations in the promoter region of the ovine MC1R gene in Shirazi Karakul sheep.

**MATERIALS AND METHODS**

Blood samples were collected from 83 Shirazi Karakul sheep breed raised at a farm in Fars Province, Iran. Genomic DNA was extracted using the AccuPrep® genomic DNA extraction kit, according to the manufacturer’s instructions. The 380 bp of promoter region (-377 to +3) of the MC1R gene was amplified through Polymerase Chain Reaction (PCR) under the following conditions: one 5 minutes denaturing cycle at 95°C, 30 cycles of 95°C for 45 seconds, 63°C for 30 seconds and 72°C for 30 seconds, along with one elongation cycle at 72°C for 10 minutes. Forward (5′-ctgcaggtccttctggactc-3′) and reverse (5′-cattgctcctcttctgagca -3′) primers were designed using Primer 3 software (http://frodo.wi.mit.edu/primer3/) according to the MC1R sequence (Z31369.1). Twenty-five µl of PCR were carried out into 0.2 ml PCR tubes, using a PCR kit with the lyophilized components (BiONEER). Each tube contained 1.0 unit of Taq DNA polymerase, 10 mM of Tris-HCl (pH 9), 30 mM of KCl, 1.5 mM of MgCl\textsubscript{2}, 250 µM of each dNTP, 20 pmol of each primer and 50 ng genomic DNA. Single-Strand Conformation Polymorphism (SSCP) was employed to genotype the promoter region of the MC1R gene. Two µL of each PCR product was mixed with 6 µl stop solution (95% formamide, 20 mM EDTA, 0.025% xylene cyanol and 0.025% Bromophenol Blue). The samples were heated for 5 minutes at 95°C.
and immediately cooled on ice. The total volume was loaded onto a 10% polyacrylamide gel (37.5:1 Acrylamide/Bisacrylamide). Electrophoresis was carried out at room temperature in 0.5X TBE buffer for 20 hours. The gels were subsequently fixed in 10% acetic acid, stained with 0.15% AgNO3 and revealed with 1.5% Na2CO3.

The purified PCR products were commercially sequenced and multiple sequence alignment performed with CLUSTALW (http://workbench.sdsc.edu). The sequences for promoter region of MC1R gene were submitted to NCBI (JF800648, JF800649, JF800650 and JF800651). In silico analysis of novel identified mutations within the MC1R promoter was performed using TFSEARCH v1.3 (www.cbrc.jp/research/db/TFSEARCH.html) and MatInspector professional 8.2 from Genomatix Software Complex (http://www.genomatix.de), and compared with the wild-type sequence to identify any association regarding transcription factor binding affinity.

The analysis of the promoter region through PCR-SSCP revealed four conformational patterns in the amplified fragment of the population analyzed (Figure 1). The frequencies of SSCP patterns, obtained through directed counting, of A, B, C, and D patterns for the screened fragment were 0.024, 0.036, 0.265 and 0.675, respectively.

DNA sequences of amplicons representing individual SSCP patterns revealed that the D variant (access number JF800651) matched the sequence Z31369.1 previously published (Barrett et al., 1994), while the A, B and C variants corresponded with new DNA sequences JF80050, JF80049 and JF80048 respectively. A comparison of the newly obtained MC1R sequences throughout the present study with the published MC1R sequence (Z31369.1) through CLUSTALW revealed three novel single base substitutions and a novel insertion (Figure 2).

The bioinformatics analysis of the promoter sequence using TFSEARCH and

Figure 1. Four different SSCP patterns from an amplified fragment of the MC1R promoter region.
Figure 2. Sequence chromatographs of genetic variations in the promoter region of MC1R gene. (a) Substitution of C to A at position -89; (b) Substitution of T to G at position -100; (c) Substitution of G to A at position -206, and (d) An insertion of 26-nucleotide at position -126 relative to the start codon.

MatInspector programs predicted no consensus TATA-box motif at an appropriate position, suggesting that the sheep MC1R gene might be among the TATA-less genes. Instead, multiple putative transcription factor binding sites for Ets, AML-1a (two binding sites), NF-E2, MZF1, USF, Oct-1 and GATA-1 were detected in the promoter region.

In silico analysis of the 26-nucleotide insertion polymorphism on potential cis-acting elements predicted that the insertion of the 26 nucleotides (TCCAGGATCCAGGATCAAGGACCCA G) would not disrupt the consensus sequences for any known transcription factor, but instead would create a putative binding site for MOK2 (Figure 3).

DISCUSSION

Throughout the present study, considerable variation was found to exist in the promoter region of MC1R gene as revealed by 4 different SSCP variants which further identified 3 Single Nucleotide Polymorphic sites (SNPs) and an insertion of 26 nucleotides.

In silico analysis revealed a C> A sequence variation at position -89 located within the core sequence of a USF binding site. The analysis of mutated sequence also revealed that this C> A substitution abolished the binding site of USF (Table 1). The USF transcription factors, including ubiquitously expressed USF1 and USF2, belong to the family of C-terminal basic Helix–Loop–Helix leucine Zipper (bHLH-Zip) proteins that bind to E-box elements containing the core sequence "CACCTG" (Grandori et al., 2000). Nucleotide sequence analysis of the humane MC1R promoter region revealed that the E-box is present immediately upstream of the transcriptional initiation site (Carr and Sharp, 1990). Mutation of the CA dinucleotide to TC at E-box in the promoter region eliminated USF binding site of plasminogen activator inhibitor-1 gene (Kutz et al., 2006).

Transcriptional regulation by USF contextually dependent in that activation may require expression of specific co-activators, or cooperative interactions with other transcription factors (Qyang et al., 1999). Additionally, in some promoter regions USF has a repressive effect on transcription (Harris et al., 2000, Naghavi et al., 2001). In vitro studies have
Genetic Variation of MC1R Promoter Region

**Figure 3.** Sequence of the MC1R-promoter, identified genetic variations (red, bold) and potential transcription factor binding sites (grey boxes). The parenthesis shows the 26 nucleotide insertion at position -126 relative to the start codon.

**Table 1.** Transcription factors within the promoter region of **MC1R** gene in the wild-type and mutated sequences.

<table>
<thead>
<tr>
<th>Transcription factor</th>
<th>Nucleotide position</th>
<th>Sequence motifs</th>
<th>Wild-type matrix match</th>
<th>Mutated matrix match</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ets</td>
<td>-349 to 337</td>
<td>actcaggaatagt</td>
<td>87.0</td>
<td>87.0</td>
</tr>
<tr>
<td>AML-1α</td>
<td>-208 to 204</td>
<td>tgggg</td>
<td>87.4</td>
<td>88.7</td>
</tr>
<tr>
<td>NF-E2</td>
<td>-197 to -187</td>
<td>atgactctgtct</td>
<td>85.2</td>
<td>85.2</td>
</tr>
<tr>
<td>MZF1</td>
<td>-171 to -164</td>
<td>ggagggaga</td>
<td>93.0</td>
<td>93.0</td>
</tr>
<tr>
<td>MOK2</td>
<td>-136 to -133</td>
<td>aagg</td>
<td>-</td>
<td>98.1</td>
</tr>
<tr>
<td>AML-1α</td>
<td>-102 to -98</td>
<td>tgtgg</td>
<td>1.00</td>
<td>87.4</td>
</tr>
<tr>
<td>USF</td>
<td>-90 to -83</td>
<td>cccagtcg</td>
<td>90.8</td>
<td>Abolished</td>
</tr>
<tr>
<td>Oct-1</td>
<td>-85 to -72</td>
<td>tgcatgtacctgt</td>
<td>85.7</td>
<td>85.7</td>
</tr>
<tr>
<td>GATA-1</td>
<td>-55 to -46</td>
<td>ctggagtaacg</td>
<td>86.1</td>
<td>86.1</td>
</tr>
</tbody>
</table>

...demonstrated a possible role for USF-1 and USF-2 in controlling cell proliferation and transformation by competing with oncogenic factor c-MYC for the binding of target promoters (Luo et al., 1996). Targeted disruption of USF binding site was responsible for a 50% reduction of transcriptional activity (Meccia et al., 2003).

The **T** to **G** substitution at position -100 in this study is located within an AML-1α binding motif. Analysis revealed that the matrix match score of wild-type sequence is higher than the mutated sequence (Table 1), indicating that AML-1α transcription factor possesses a higher affinity to the -100T variant as compared with the -100G variant.
The AML1 transcription factor binds to the consensus DNA sequence \textit{TGT/cGGT}, which is present in a number of promoters and enhancers of viral as well as cellular genes (Bae \textit{et al.}, 1993). The AML1 is normally expressed in all lineages of hematopoietic cells acting as a regulator of the expression of various genes, specific to hematopoiesis (Ito \textit{et al.}, 1997). Mutation of AML-1-sites at the promoter region of human \textit{BPI} gene decreased promoter activity in myeloid cells (Lennartsson \textit{et al.}, 2003).

The significant finding of this work is the identification of a novel insertion/deletion (Indels) of 26-nucleotides at position -126 relative to the start codon in the \textit{MC1R} promoter (Figure 2). Currently, there is an increasing focus on polymorphisms of the short insertion and deletion types in genomic research related of different organisms (Bhangale \textit{et al.}, 2005; Ometto \textit{et al.}, 2005; Brandström \textit{et al.}, 2007). Indels have been recognized as an abundant source of genetic markers that are widely spread across the genome, though they are not as common as SNPs (Mills \textit{et al.}, 2006). Several studies have reported the relationship between indels with a number of diseases in human. For example, an indel (94 Insertion= Deletion ATTG) in the promoter of human \textit{NFKB1} (nuclear factor-kB protein complex) was related to carcinogenic processes in renal cell carcinoma, colorectal carcinoma, chronic lymphocytic leukemia (Riemann \textit{et al.}, 2006) and in superficial bladder cancer (Riemann \textit{et al.}, 2007; Tang \textit{et al.}, 2010). A functional indel within intron 16 of the \textit{ACE} gene (Angiotensin-converting enzyme) was suggested to be associated with panic disorder in a potentially gender-specific way (Olsson \textit{et al.}, 2004), whole-body insulin sensitivity and impaired glucose tolerance (Bonnet \textit{et al.}, 2008). In these cases, the indel itself may play either a direct role in regulation of gene transcription or may be in linkage disequilibrium with regulatory elements of the gene.

\textit{In silico} analysis, using MatInspector revealed that the 26-nucleotide insertion created a putative binding site for murine and humane MOK2 transcription factor (Figure 3). The murine and human MOK2 factors can recognize both DNA and RNA through their zinc finger motifs (Arranz \textit{et al.}, 1997). This dual affinity of MOK2 for RNA and DNA suggests that MOK2 may be involved in transcription and posttranscriptional regulation of MOK2 target genes. There is strong evidence that the MOK2 transcription factor can bind to the 8bp present in the promoter region of the humane \textit{IRBP} (Interphotoreceptor Retinoid-Binding Protein) gene and repress transcription by competing with the CRX activator for DNA binding (Dreuillet \textit{et al.}, 2002; Arranz \textit{et al.}, 2001). A promoter assay of human \textit{MC1R} revealed that a deletion from -517 to -447 (location of the SP-1 site) completely abolishes promoter activity (Morö \textit{et al.}, 1999). The sub-lethal gene, lethal roan, is present in the Karakul breed known in Russia as Shirazi, important in the production of grey pelts for the fur trade. It is now known that the gene that causes the grey color is sub-lethal when homozygous, which explains why grey to grey mating of Karakul sheep is avoided (Lundie, 2011). Gene expression studies could help find the association between 26 indel in the \textit{MC1R} promoter and risk of lethality in grey color of Karakul sheep.

**CONCLUSIONS**

In summary, the promoter region of \textit{MC1R} sequence bears the features of a TATA-fewer genes and contains consensus binding sites for several transcription factors, including Ets, AML-1a, NF-E2, MZF1, USF, Oct-1 and GATA-1, that may be functionally important. It is believed that is the first study of SNP identification in the promoter region of the ovine \textit{MC1R} gene that could up-regulate gene expression. Gene expression analysis is needed to identify the functional role of these identified polymorphic sites in the promoter region of the ovine \textit{MC1R} gene that we identified.
ACKNOWLEDGEMENTS

This work was supported by Yasouj University (Project No. 1381), for which the authors are grateful to.

REFERENCES


بررسی توالی ناحیه پرموتر زن ملانوکورتین-1 در گوسفند نژاد فوق گل

م. محقق دولت آبادی, ج. حیبی زاد

چکیده
گیرنده ملانوکورتین-1 (MC1R) توسط چاپیگاه گسترش کد شده و نقش مهمی در تعیین رنگ پوست گوسفند نژاد فوق گل است. در این مطالعه، با استفاده از تکنیک تفاوت فرم فضایی رشته‌های ضد مورد بررسی قرار دادیم. در کل تعداد 42 نمونه توالی MC1R ناحیه پرموتر زن DNA مشاهده شد که پس از تجزیه و تحلیل توالی 3 چاپیگاه جدید شکلی تک نوکلئوتیدی و یک جدید SSCP
شکلی اضافه و حذف شده‌گی در جمعیت مورد بررسی آشکار شد. تجزیه و تحلیل انیسیلوکو ناحیهٔ پروموتر زن MC1R چابگاهی برای جوجه TATA در چابگاه مناسب پیش بینی نکرد اما چابگاه های Oct-1, USF, MZF1, NF-E2, AML-1, ETs, USF متعددی برای فاکتور های روتوپی و شناسایی کرد. تجزیه و تحلیل چابگاه های جهت شکلی شناسایی شده نشان داد که چند USF شکلی چابگاه 89- قبل از کدون آغازین چابگاه اتصال فاکتور روتوپی 1- با مراقبت می‌کرد. USF شکلی شناسایی شده در چابگاه 100- داخل چابگاه اتصال فاکتور روتوپی 1- قرار گرفته است. اضافه شدن 26 تولید در چابگاه 126- چابگاه اتصالی برای فاکتور روتوپی MOK2 ابجا می‌کند. نقص ممکن تنوع انتی‌تیک شناسایی شده می‌تواند با استفاده از تجزیه و تحلیل بین چهارمورد تایید قرار گیرد.