**Abstract**

The adrenergic receptor kinase 1 (ADRBK1), thromboxane A2 receptor (TBXA2R) and vascular endothelial growth factor (VEGFA) regulatory (r) single nucleotide polymorphisms (SNPs) found in the potential stimulating protein-1 (SP1) and Kruppel-like factor-4 (KLF4) transcriptional factor binding sites (TFBS) within these genes are in linkage disequilibrium (LD). The LD may result from rSNP alleles that create TFBS for the KLF4 and SP1 transcriptional factors (TF) since the alternate rSNP alleles do not create these TFBS. Consequently, haplotypes carrying the rSNP alleles that create KLF4 and SP1 TFBS are essential for ADRBK1, TBXA2R and VEGFA gene regulation by these TFs.

**Keywords**

ADRBD1 (GRK2), TBXA2R, VEGFA, KLF4, SP1, TFBS, rSNP, LD

---

1. **Introduction**

The transcription factor (TF) stimulating protein-1 (SP1) participates in the transcription regulation of Kruppel-like factor-4 (KLF4), a zinc finger-containing TF and together they are part of a (SP1/KLF) family of TFs which play a role in diverse cellular processes, including vascular smooth muscle cell (VSMC) proliferation, cell differentiation, apoptosis and, oncogenic processes [1] [2], induction of pluripotent stem cells [3] and control of gene transcription [4]-[7]. There have been at least 20 KLFs identified in mammals with each individually participating in one of the above biological functions [6]. KLF4 has been shown to play a key role in pathological vascular processes and is considered as a molecular switch in regulating VSMC function [6]. KLF4 and SP1 physically interact, occupy the angiotensin II type 1 receptor (AT1R) promoter and induce AT1R transcription in...
VSMCs in a co-operative manner under basal conditions [8]. KLF4 and SP1 also have transcriptional factor binding sites (TFBS) in the regulatory regions of the adrenergic receptor kinase 1 (ADRBK1), thromboxane A2 receptor (TBXA2R) and vascular endothelial growth factor (VEGFA) genes. The SP1 TF has been shown to bind its transcriptional factor binding site (TFBS) and regulate the ADRBK1 [9], TBXA2R [10] [11] and VEGFA [12] genes. The ADRBK1 gene (also known as GRK2) is encoded on chromosome 11 and is expressed as a ubiquitous cytosolic enzyme that specifically phosphorylates the activated form of the beta-adrenergic and related G-protein-coupled receptors. Abnormal coupling of beta-adrenergic receptor to G protein is involved in the pathogenesis of the failing heart. The TBXA2R gene is encoded on chromosome 19 and is a member of the G protein-coupled receptor family. The protein interacts with thromboxane A2 to induce platelet aggregation and regulate hemostasis. The VEGFA gene is encoded on chromosome 6 and is usually expressed as a 46-kDa disulfide-linked homodimer. VEGFA is a signaling protein involved in the regulation of angiogenesis, vasculogenesis and endothelial cell growth. It induces endothelial cell proliferation, promotes cell migration, inhibits apoptosis and induces permeabilization of blood vessels.

Single nucleotide changes that affect gene expression by impacting gene regulatory sequences such as promoters, enhancers, and silencers are known as regulatory SNPs (rSNPs) [13]–[16]. A rSNP within a transcriptional factor binding site (TFBS) can change a transcriptional factor’s (TF) ability to bind its TFBS [17]–[20] in which case the TF would be unable to effectively regulate its target gene [21]–[25]. This concept is examined for ADRBK1, TBXA2R and VEGFA rSNPs found in potential SP1 and KLF4 TFBS within these genes. These rSNPs within each gene have been reported to be in linkage disequilibrium (LD) [26]–[28] and this disequilibrium in conjunction with their SP1 and KLF4 TFBS were examined in this study. In this report LD is considered to be the non-random association of rSNP alleles within the gene.

2. Materials and Methods

Identifying TFBS

The JASPAR CORE database [29] [30] and ConSite [31] were used to identify the KLF and SP TFBS in this study. JASPAR is a collection of transcription factor DNA-binding preferences used for scanning genomic sequences where ConSite is a web-based tool for finding cis-regulatory elements in genomic sequences. The TFBS and SNP location within the binding sites have previously been discussed [32]. The Vector NTI Advance 11 computer program (Invitrogen, Life Technologies) was used to locate the TFBS in the ADRBK1 gene (NCBI Ref Seq NM_001619) from 12.78 kbp upstream of the transcriptional start site to 3 kbp past the 3’UTR which represents a total of 36 kbp.

The VNTI program was also used to locate TFBS in the TBXA2R gene (NCBI Ref Seq NM_201636) from 9.4 kb upstream of the transcriptional start site to 1.4 kb past the 3’UTR which represents a total of 17.1 Kbp and in the VEGFA gene (NCBI Ref Seq NM_001171626) from 2.2 kb upstream of exon one to 1.7 kb past the 3’UTR which represents a total of 19.6 Kbp. The JASPAR CORE database was also used to compute each nucleotide occurrence (%) within the TFBS where upper case lettering indicate that the nucleotide occurs 90% or greater and lower case letter less than 90%. The occurrence of each SNP allele in the TFBS is also computed from the database (Table 1).

3. Results

3.1. ADRBK1 rSNPs and TFBS

The ADRBK1 gene transcribes GRK2 a serine/threonine kinase which is a ubiquitous cytosolic enzyme that specifically phosphorylates the activated form of the beta-adrenergic and related G-protein-coupled receptors. The rs948988 SNP common ADRBK1-G allele creates three TFBS for the KLF1, KLF4 and KLF5 TFs which are involved with erythocyte development, activation and repression (Table 1). The rs4370946 SNP common ADRBK1-C allele creates TFBS for the KLF4, KLF5, SP1, and SP2 TFs which are involved with transcription activation and repression (Table 1). The minor alleles for the two rSNPs do not create any KLF or SP TFBS.

3.2. TBXA2R rSNPs and TFBS

The TBXA2R gene encodes a member of the G protein-coupled receptor family and the protein interacts with
Table 1. ADRBK1, TBX42R and VEGFA rSNPs that regulate KLF & SP TFBS. Listed are the rSNPs, gene location, allele frequency, transcriptional factors (TFs) and protein name, number of times the TFBS are found in each gene. Where the upper case nucleotide occurs at least 90% of the time in the TFBS region and red is the SNP location. Also listed is the DNA strand orientation. Below the TFBS is the rSNP allele occurrence (%) obtained from the Jaspar Core database. Note: TFs can bind to more than one nucleotide sequence.

<table>
<thead>
<tr>
<th>GENE/SNP</th>
<th>Allele</th>
<th>TFs</th>
<th>Protein name</th>
<th># of Sites</th>
<th>TFBS</th>
<th>Strand</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADRBK1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs948988 (A/G)</td>
<td>G</td>
<td>KLF1</td>
<td>Kruppel-like factor 1 (erythroid)</td>
<td>1</td>
<td>tgaCaCaCCga</td>
<td>minus</td>
</tr>
<tr>
<td>Intron 2</td>
<td>0.92</td>
<td>KLF4</td>
<td>Kruppel-like factor 4 (gut)</td>
<td>1</td>
<td>tcGGtggtg</td>
<td>plus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>KLF5</td>
<td>Kruppel-like factor 5 (intestinal)</td>
<td>1</td>
<td>cegaCtCCCCc</td>
<td>minus</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>c = 87%</td>
<td></td>
</tr>
<tr>
<td>rs4370946 (C/T)</td>
<td>C</td>
<td>KLF4</td>
<td>Kruppel-like factor 4 (gut)</td>
<td>1</td>
<td>tGGGaGgGgg</td>
<td>minus</td>
</tr>
<tr>
<td>3'UTR</td>
<td>0.92</td>
<td>KLF5</td>
<td>Kruppel-like factor 5 (intestinal)</td>
<td>1</td>
<td>cccgCtCCCa</td>
<td>plus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SP1</td>
<td>Specificity Protein 1</td>
<td>1</td>
<td>tgcCgCtccc</td>
<td>plus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SP1</td>
<td>Specificity Protein 1</td>
<td>1</td>
<td>cccgCtCCcag</td>
<td>plus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SP2</td>
<td>Specificity Protein 2</td>
<td>1</td>
<td>tgcCgCtccagt</td>
<td>plus</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>C = 100%</td>
<td></td>
</tr>
<tr>
<td>TBX42R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2238633 (A/C)</td>
<td>C</td>
<td>KLF4</td>
<td>Kruppel-like factor 4 (gut)</td>
<td>1</td>
<td>aGGGtgGgt</td>
<td>minus</td>
</tr>
<tr>
<td>Intron 1</td>
<td>0.83</td>
<td>SP1</td>
<td>Specificity Protein 1</td>
<td>2</td>
<td>CcCactCctgc</td>
<td>plus</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>c = 86%</td>
<td></td>
</tr>
<tr>
<td>rs4523 (A/G)</td>
<td>T</td>
<td>KLF4</td>
<td>Kruppel-like factor 4 (gut)</td>
<td>1</td>
<td>tGGGtgGata</td>
<td>minus</td>
</tr>
<tr>
<td>e.924T&gt;C</td>
<td>0.45</td>
<td></td>
<td></td>
<td></td>
<td>t = 7%</td>
<td></td>
</tr>
<tr>
<td>Exon 3</td>
<td></td>
<td>SP1</td>
<td>Specificity Protein 1</td>
<td>1</td>
<td>cccgGtCGgtt</td>
<td>plus</td>
</tr>
<tr>
<td>rs5756 (C/T)</td>
<td>C</td>
<td>KLF4</td>
<td>Kruppel-like factor 4 (gut)</td>
<td>4</td>
<td>eGGGeGggt</td>
<td>plus</td>
</tr>
<tr>
<td>3'UTR</td>
<td>0.63</td>
<td></td>
<td></td>
<td></td>
<td>G = 98%</td>
<td></td>
</tr>
<tr>
<td>VEGFA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1570360 (A/G)</td>
<td>G</td>
<td>KLF4</td>
<td>Kruppel-like factor 4 (gut)</td>
<td>2</td>
<td>gtGgaGgGgc</td>
<td>plus</td>
</tr>
<tr>
<td>Promoter</td>
<td>0.78</td>
<td>SP1</td>
<td>Specificity Protein 1</td>
<td>1</td>
<td>CCCtTCCACA</td>
<td>minus</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>c = 77%</td>
<td></td>
</tr>
<tr>
<td>rs3435723 (D/I)</td>
<td>I</td>
<td>KLF5</td>
<td>Kruppel-like factor 5 (intestinal)</td>
<td>1</td>
<td>gtcccCaCtCt</td>
<td>plus</td>
</tr>
<tr>
<td>Promoter</td>
<td>0.28</td>
<td>KLF5</td>
<td>Kruppel-like factor 5 (intestinal)</td>
<td>1</td>
<td>acetCtCCCa</td>
<td>plus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SP1</td>
<td>Specificity Protein 1</td>
<td>1</td>
<td>gtcCtCACtct</td>
<td>plus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SP1</td>
<td>Specificity Protein 1</td>
<td>1</td>
<td>actCtCtCCc</td>
<td>plus</td>
</tr>
</tbody>
</table>

(I)ntertation = 18bp (GGTCCCACCTCTTCCCCACA)
thromboxane A2 to induce platelet aggregation and regulate hemostasis. The activity of this receptor is mediated by a G-protein that activates a phosphatidylinositol-calcium second messenger system. The rs2238633 SNP \( TBA2R-C \) allele in intron one creates two TFBS for the KLF4 and SP1 TFs which are involved in transcriptional regulation and repression. The rs4523 SNP \( TBA2R-T \) allele in exon three and the rs5756 SNP \( TBA2R-C \) allele in the 3’UTR each create one TFBS for the KLF4 TF. The other alleles for the three rSNPs do not create any KLF or SP TFBS.

3.3. \( VEGFA \) rSNPs and TFBS

The \( VEGFA \) gene is a member of the PDGF/VEGF growth factor family and encodes a protein that is often found as a disulfide linked homodimer. This protein is a glycosylated mitogen that specifically acts on endothelial cells and has various effects, including mediating increased vascular permeability, inducing angiogenesis, vasculogenesis and endothelial cell growth, promoting cell migration, and inhibiting apoptosis. The rs1570360 SNP \( VEGFA-G \) allele in the promoter at \(-1190 \) bp from the transcriptional start site (TSS) creates two TFBS for the KLF4 and SP1 TFs each in the same location but opposite strands of the DNA duplex (Table 1). The rs34357231 SNP is a 18bp insertion/deletion (I/D) polymorphism located in the promoter at \(-2549 \) bp from the TSS [33]. The rs34357231 SNP \( VEGFA-I \) allele creates TFBS for the KLF5 and SP1 TFs. The other alleles for the two rSNPs do not create any KLF or SP TFBS.

4. Discussion

Genome-wide association studies (GWAS) over the last decade have identified nearly 6500 disease or trait-predisposing SNPs where only 7% of these are located in protein-coding regions of the genome [34] [35] and the remaining 93% are located within non-coding areas [36] [37] such as regulatory or intergenic regions. SNPs which occur in the putative regulatory region of a gene where a single base change in the DNA sequence of a potential TFBS may affect the process of gene expression are drawing more attention [13] [15] [38]. A SNP in a TFBS can have multiple consequences. Often the SNP does not change the TFBS interaction nor does it alter gene expression since a transcriptional factor (TF) will usually recognize a number of different binding sites in the gene. In some cases the SNP may increase or decrease the TF binding which results in allele-specific gene expression. In rare cases, a SNP may eliminate the natural binding site or generate a new binding site. In which cases the gene is no longer regulated by the original TF. Therefore, functional rSNPs in TFBS may result in differences in gene expression, phenotypes and susceptibility to environmental exposure [38]. Examples of rSNPs associated with disease susceptibility are numerous and several reviews have been published [38]-[41].

In this study the rs948988 rSNP \( ADRBK1-G \) allele [G (+ strand) or C (−strand)] located in the KLF1, KLF4 and KLF5 TFBS have a 100%, 98% and 87% occurrence, respectively, and consequently have been reasonably well conserved in human evolution (Table 1). Each of the three potential KLF TFBS occur only one time in the gene (Table 1) and therefore the rSNP would probably have a great impact on the KLF TFs regulating the gene. The rs4370946 rSNP \( ADRBK1-C \) allele [C (+ strand) or G (−strand)] located in the two KLF and three SP TFBS have all been well conserved in human evolution ranging from 77 to 100% in occurrence (Table 1).

The rs2238633 rSNP \( TBA2R-C \) allele [C (+ strand) or G (−strand)] located in the KLF4 and SP1 TFBS have a 90 and 86% occurrence, respectively, and have been well conserved in human evolution (Table 1). The KLF4 TFBS occurs once while the SP1 TFBS occurs twice in the gene. The rs4523 rSNP \( TBA2R-T \) allele [T (-strand)] located in a KLF4 TFBS has a 7% occurrence and not well conserved in human evolution, therefore, probably would not have much impact on regulating the gene (Table 1). The rs5756 rSNP \( TBA2R-C \) allele [C (−strand) or G (+strand)] located in a KLF4 TFBS has a 98% occurrence in humans and this TFBS occurs four times in the gene (Table 1). The rs1570360 SNP \( VEGFA-G \) allele [G (+strand) or C (−strand)] is a well conserved nucleotide (98%) in the TFBS for the KLF4 TF and moderately conserved (77%) in the TFBS for the SP1 TF. These potential TFBS occur twice and once in the gene, respectively. The rs34357231 SNP \( VEGFA-I \) allele has an insertion frequency of 0.28. All of the potential KLF5 and SP1 TFBS are created by the insertion and do not exist when the 18bp sequence is deleted.

When examining the table, it can be seen from the KLF4/SP1 TFBS representing the \( ADRBK1 \) (rs4370946), \( TBA2R \) (rs2238633) and \( VEGFA \) (rs1570360) rSNPs that the KLF4 TF binds the duplex DNA on one strand while SP1 binds the DNA at the same location on the opposite strand which suggests that these two TFs may be involved in duplex DNA strand separation [42] [43] during transcription. Since these TFs potentially bind the
DNA at the same location, a nucleotide change in their respective TFBS could affect their regulation of the respective genes. 

LD between SNPs in the regulatory region of a gene can indicate strong associations of certain haplotypes with sickness or disease [28]. Consequently, an evaluation was made within the ADRBK1 [9], TBXA2R [10] [11] and VEGFA [12] genes for the rSNPs in LD [26] and their effect on potential KLF/SP TFBS. Since it has been shown that KLF4 and SP1 physically interact, occupy the angiotensin II type 1 receptor (AT1R) promoter and induce AT1R transcription in VSMCs in a co-operative manner under basal conditions [8], the evaluation was based on these two TFs. The LD between the ADRBK1 rs948933 and rs437046 rSNPs may result from the lack of a SP1 TFBS regulated by rs948988 in intron two while there is a SP1 TFBS regulated by rs4370946 in the 3' UTR of the gene (Table 1). The LD between the TBXA2R rs2238633, rs4523 and rs5756 rSNPs may also result from the lack of a SP1 TFBS regulated by the rs4523 rSNP in exon three and by the rs5756 rSNP in the 3'UTR of the gene (Table 1). The LD between the VEGFA rs1570360 and rs34357231 rSNPs in the promoter may result from the lack of a KLF4 TFBS in the insertion region created by the rs34357231 rSNP. For KLF4 and SP1 TF interaction, there is another KLF4 TFBS upstream regulated by the rs1570360 rSNP. Since the alternate alleles of the above rSNPs do not create any KLF or SP TFBS, it can be assumed that haplotypes carrying rSNP alleles which create KLF and SP TFBS are essential in the regulation of these genes and could be responsible for the LD.

5. Conclusion

Single nucleotide changes in TF motifs have the ability to alter gene regulation and thereby result in disease. There are many reports appearing in the literature describing human disease and the association with SNPs [44]; however, few of these reports are examining the SNP location for accompanying changes in TFBS that would affect gene regulation [45]. In the present study the SNPs in the regulatory region of the ADRBK1, TBXA2R and VEGFA genes can alter the KLF and SP1 binding sites rendering these TFs unable to regulate these genes.

Dedication

This manuscript is dedicated to the memory of Richard M. Buroker.

References


AKT3, ANGPTL4, eNOS3, and VEGFA Associations with High Altitude Sickness in Han and Tibetan Chinese at the Qinghai-Tibetan Plateau.


