Journal of Clinical & Medical Genomics

Buroker et al., J Clin Med Genom 2015, 3:1
DOI: 10.4172/2472-128X.1000120

Research Article Open Access

SNPs, Linkage Disequilibrium and Transcriptional Factor Binding Sites Associated with Acute Mountain Sickness among Han Chinese at the Qinghai-Tibetan Plateau

Norman E Buroker[1]*, Xue-Han Ning[1,2], Kui Li[3], Zhao-Nian Zhou[4], Wei-Jun Cen[3], Xiu-Feng Wu[4], Wei-Zhong Zhu[5], C Ronald Scott[1] and Shi-Han Chen[1]

[1]*Department of Pediatrics, University of Washington, Seattle, WA USA*
[2]*Division of Cardiology, Seattle Children's Hospital Institute Foundation, Seattle, WA USA*
[3]*Lhasa People Hospital, Tibet*
[4]*Laboratory of Hypoxia Physiology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, China*
[5]*Center for Cardiovascular Biology and Regenerative Medicine, University of Washington, Seattle, WA USA*

Abstract

Acute mountain sickness (AMS) occurs in up to 50% of individuals ascending to high altitudes greater than 2600 meters. An AMS Han Chinese and a normal Han group were compared for 17 simple nucleotide polymorphisms (SNPs) within 9 genes that have been associated with AMS. The SNPs were analyzed with respect to linkage disequilibrium (LD) between intra- and intergenic SNP alleles and alterations in transcriptional factor binding sites (TFBS). Included in the study was the angiotensin-converting enzyme (ACE) (rs4340), the angiotensinogen (AGT) (rs699) and the angotensin II type 1 receptor (AGTR1) (rs5186) SNPs from the renin-angiotension system (RAS) as well as the GNB3 (rs2071057) SNP from G-protein signaling and a LDL apolipoprotein B (APOB) (rs693) SNP. The endothetal Per-Arnt-Sim (PAS) domain protein 1 (EPAS1) SNP and two egl nine homolog 1 (EGLN1) SNPS (rs480902 and rs516651) from the hypoxia-inducible factor (HIF) oxygen signaling pathway were included. SNPs analyzed in the vascular endothelial growth factor (VEGF) signaling pathway are the v-akt murine thymoma viral oncogene homolog 3 (AKT3) (rs4590656 and rs2291409), the endothelial cell nitric oxide synthase 3 (eNOS3) (rs1007311 and rs1799983) and the (VEGFA) (rs79469752, rs13207351, rs28357093, rs1570360 and rs3025039). These SNP alleles alter the TFBS for TF binding. Pair-wise LD was computed between SNPs. An increase in LD occurred in 32 pair-wise comparisons while a decrease was found in 22 pair-wise comparisons between the AMS and controls. Increases and decreases in LD pairs were found within and between signaling pathways and systems indicating the interaction of SNP alleles or potential TFBS from different areas of the genome. The most drastic change in TFBS occurs with ACE (I/D) SNP (rs4340) where the ACE-I allele generates 84 potential TFBS while the ACE-D allele generates only four binding sites. The alteration in TFBS generated by the 17 SNPs is discussed with respect to AMS.

Keywords: Acute mountain sickness; LD; SNPs; TFBS

Introduction

Acute mountain sickness (AMS) is very common among some individuals who ascend to altitudes greater than 2600 m. The illness is characterized by headache, lightheadedness, breathlessness, fatigue, insomnia, anorexia, and nausea [1,2]. The symptoms of the illness begin 2-3 hours after ascent to the higher altitude. The medical condition is generally self-limiting in the individual where most symptoms disappear after two or three days, although insomnia can persist longer [3]. Individuals with the sickness will enter the hospital to be treated under emergency conditions. The illness will resolve itself if no additional altitude is attempted; however, in some cases the descent to a lower altitude is necessary in order to reverse the condition. The precise pathogenesis of AMS is not well understood, but hypoxia is considered to be the major factor [4-7], which raises the question of why some individuals are susceptible to the sickness while others are not, under the same high altitude environment.In an effort to provide in cite into this question, we have previously published on known genetic associations of simple nucleotide polymorphisms (SNPs) with high altitude sickness [8-11]. In the present study, we examine the interaction of seventeen SNPs in nine genes [8-11] with regard to AMS and the effect these SNPs have on potential changes in transcriptional factor binding sites (TFBS).

Included in the study is the renin-angiotensin system (RAS) and G protein signaling which are related to hypertension [12] as well as polymorphisms affecting blood levels of low-density lipoprotein (LDL) and triglyceride [13,14]. The RAS plays a major role in the regulation of systemic arterial blood pressure and is also involved in the regulation of pulmonary vascular tone. We included three RAS gene polymorphisms, the angiotensin-converting enzyme (ACE) insertion/deletion (rs4340), the angiotensinogen (*AGT*) M268T (rs699) and the angotensin II type 1 receptor (*AGTR1*) A1166C (rs5186) variants. The *GNB3* gene encodes the Gβ3 subunit of heterotrimeric G proteins (α, β, γ), which are key components of intracellular signal transduction present in all cells of the body [15] and has been associated with hypertension [16]. We included the *GNB3*, A(-350)G (rs2071057) SNP in the promoter region of the gene, which is in linkage disequilibrium with two other *GNB3* SNPs within the gene, C825T (exon 10) and C1429T (exon 11) [17]. Since LDL cholesterol and triglyceride concentrations are strongly influenced by the genetic constitution of each individual and physical activity has

*** Corresponding author:** Norman E. Buroker, Department of Pediatrics, University of Washington, Seattle, WA USA, Tel: 206 616 0472; E-mail: nburoker@u.washington.edu

Received March 17, 2015; **Accepted** April 04, 2015; **Published** April 12, 2015

Citation: Buroker NE, Ning XH, Li K, Zhou ZN, Cen WJ, et al, (2015) SNPs, Linkage Disequilibrium and Transcriptional Factor Binding Sites Associated with Acute Mountain Sickness among Han Chinese at the Qinghai-Tibetan Plateau J Clin Med Genom 3: 120. doi: 10.4172/2472-128X.1000120

a role in determining an individual's lipid profiles, we included a LDL apolipoprotein B (*APOB*) (rs693) SNP in the coding region of the gene [8].

Two genes from the hypoxia-inducible factor (HIF) oxygen signaling pathwayare included in the study [10]. The endothetal Per-Arnt-Sim (PAS) domain protein 1 (*EPAS1*) gene, which encodes hypoxia-inducible-factor-2 alpha (*HIF2A*) a transcription factor that responses to hypoxia conditions. The *EPAS1* gene has a SNP [ch2: 46441523 (hg18)] in intron five located five base pairs from the beginning of exon six that has a 78% frequency difference between Tibetan and Han Chinese [18]. The egl nine homolog 1 (*EGLN1*) gene acts as a key oxygen sensor which negatively regulates the activity of the hypoxia-inducible factor-1 alpha (*HIF-1A*). Hypoxia causes an inactivation of the *EGLN1* gene thereby increasing HIF activity that induces the expression of genes which mediates the adaptive responses through glycolytic enzymes, hemeoxygenase, vascular endothelial growth factor and erythropoietin [19]. The two *EGLN1* SNPs (rs480902 and rs516651) included in the study are located in intron 2 of *EGLN1* have been associated with high altitude adaptation in human populations [10,19-22].

We also included SNPs from three genes in the vascular endothelial growth factor (VEGF) signaling pathway[9] where the VEGFA protein is a growth factor activator for angiogenesis, vasculogenesis and endothelial cell growth. Four *VEGFA* SNPs (rs79469752, rs13207351, rs28357093 and rs1570360) are found in the promoter region while a fifth *VEGFA* SNP (rs3025039) is located in the 3'UTR region. Also in this pathway is the v-akt murine thymoma viral oncogene (homolog 3) (*AKT3*) gene whose protein is a serine/threonine kinase that plays a key role in regulating cell survival, insulin signaling, angiogenesis and tumor formation. The two *AKT3* SNPs (rs4590656 and rs2291409) used in the study are located in introns. A third gene in the pathway is the endothelial cell nitric oxide synthase 3 (*eNOS3*) which produces nitric oxide (NO) and is implicated in vascular smooth muscle relaxation. NO mediates VEGF-induced angiogenesis in coronary vessels and promotes blood clotting through the activation of platelets. The two *eNOS3* SNPs (rs1007311 and rs1799983) used in the study are located in an intron and an exon, respectively.

In genetics, linkage disequilibrium (LD) is defined as the non-random association in a given population between the alleles of two or more loci [23]. LD between SNPs in the regulatory region of a gene can be used as a method of identifying associations of certain haplotypes that lead to sickness or disease in a population [9,11].This can be achieved when levels of LD between SNPs within haplotypes are seen to change substantially in a disease or sickness group when compared to the normal baseline population. In such cases, the relationship between LD, SNPs and TFBS can be used to identify potential binding changes for TFs responsible for gene regulation. Such TFBS changes could result in disease or sickness [11,24-29]. In this report, LD is considered to be the non-random association of SNP alleles within and between genes. LD was computed among the 17 SNPs and compared between the AMS and control Han Chinese group. Since these SNPs are located in potential TFBS, any changes in LD between the groups is discussed with relation to the elimination, change or addition of a punitive TFBS created by each SNP.

Materials and Methods

Study Groups

The Han Chinese who are considered upward migrants from low altitudes were used as our study source. All AMS patients in this study had been hospitalized and diagnosed at the Lhasa People Hospital (Tibet, China at 3,670 M above sea level) between 2002 and 2008. AMS was diagnosed by using the current consensus of mountain sickness in Tibet (Diagnosis and Therapeutics for Mountain Sickness, Xizang Autonomous Region), which is in accord with the Lake Louise scoring system [30]. We sampled Han AMS patients from the hospital with symptoms of acute pulmonary edema as diagnosed by a cough accompanied with pink frothy sputum. Moist or bubbling rales in the lungs was suggestive of high altitude pulmonary edema (HAPE), showing a characteristic shadow on chest X-rays. In addition to the characteristic symptoms of severe acute mountain response, acute high altitude cerebral edema (HACE) was diagnosed by ataxia, disturbance of consciousness or coma, abnormal plantar reflexes and papilledema. The AMS Han patients had recently arrived from the low land and acquired the illness within two days after reaching the high altitude of Tibet. Patients with other diseases having similar clinical manifestations were excluded. Healthy Han Chinese from the Lhasa area was randomly selected to serve as control subjects. All patients and controls sampled in the study signed an informed consent approved by the Human Ethics Committee of the Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences.

Sampling

Buccal brush samples were collected from 85 Han Chinese with AMS during the high occurrence period (spring and winter) at the Lhasa People Hospital. Samples for controls were also collected via buccal brush from unaffected Han individuals determined to be in good health upon physical examination by doctors at the Lhasa People Hospital. The controls consisted of 79 Han lowlanders who had travelled to the high altitude of Tibet. The people from the control group had lived at the 3600 M altitude for at least six months prior to being sampled. These sampling procedures have been previously described [8-10,31].

Genotyping

Genomic DNA was extracted from the buccal brushes using the PureGene DNA method from Gentra Systems, Inc., Minneapolis, MN. The DNA yield in this study ranged from 0.5 to 7.6 μg per buccal brush. We found the yield adequate for all PCR reactions conduced in the study. The Vector NTI Advance 11 computer program from Invitrogen, Carlsbad, CA, was used to develop the primers for genotyping each SNP. The genotyping methods and genotyping for each SNP have been previously outlined and reported [8-10,31]. The SNPs and genes used in the present study are found in (Table 1).

Statistical Analysis

All statistical methods used to genotype samples have been previously discussed [8-10,31]. Linkage disequilibrium (LD) [32] was computed between SNPs for the AMS and Han Chinese control groups (Figure1). The degree of genetic linkage between the 17 SNPs in each study group was estimated as Lewontin's coefficient |D'|, where no color (|D'|=0) indicates that LD is weak or non-existent and the dark red (|D'|=1) indicates that there exists strong pairwise linkage disequilibrium between SNPs (Figure 1). The change in LD between control and AMS groups among all SNPs is tabulated in (Table 2).

Transcriptional factor binding sites

The JASPAR CORE database [33,34] and ConSite [35] were used to identify the TFBS in this study. JASPAR is a collection of transcription factor DNA-binding preferencesused for scanning genomic sequences where ConSite is a web-based tool for finding cis-regulatory elements

Protein and GeneSymbol	Chromosome	SNP	SNP Location	Mutation	LD Identity
Angiotensin I Converting Enzyme (*ACE*)	17q23.3	rs4340	intron16	288bp Indel/ (ALU)	ACE
Angiotensinogen (*AGT*)	1q43.2	rs699	exon1	c.803T>C, p. Met268Thr	AGT
Angiotensin II Receptor, Type1 (*AGTR1*)	3q24	rs5186	3'UTR	c.*86A>C	AGTR1
v-akt murine thymoma viral oncogene homolog 3 (*AKT3*)	1q44	rs4590656	intron1	c.46+3654C>T	AKT3-45
v-akt murine thymoma viral oncogene homolog 3 (*AKT3*)		rs2291409	intron8	c.819+4031G>A	AKT3-22
Apolipoprotein (*APOB*)	2p24.1	rs693	exon26	c.7545C>T, p.Thr2515Thr	APOB
Endothelial PAS domain protein1 (*EPAS1*)	2p21	unknown	intron5	c.657-5C>G	EPAS1
Egl-9 familyhypoxia-inducible factor 1 (*EGLN1*)	1q42.2	rs480902	intron1	c.892-21782T>C	EGLN1-48
Egl-9 familyhypoxia-inducible factor 1 (*EGLN1*)		rs516651	intron1	c.891+14088C>T	EGLN1-51
Guanin enucleotide binding protein (Gprotein), beta polypeptide3 (*GNB3*)	12p13	rs2071057	5'UTR	c.-176G>A	GNB3
Nitric oxide synthase3 (*eNOS3*)	7q36.1	rs1007311	intron6	c.817-26A>G	eNOS3-10
Nitric oxide synthase3 (*eNOS3*)		rs1799983	exon7	c.894T>G, p.Asp298Glu	eNOS3-17
Vascular endothelial growth factor A (*VEGFA*)	6p21.1	rs79469752	promoter	c.-663C>T	VEGFA-79
Vascular endothelial growth factor A (*VEGFA*)		rs13207351	promoter	c.-650A>G	VEGFA-13
Vascular endothelial growth factor A (*VEGFA*)		rs28357093	promoter	c.-639A>C	VEGFA-28
Vascular endothelial growth factor A (*VEGFA*)		rs1570360	promoter	c.-614A>G	VEGFA-15
Vascular endothelial growth factor A (*VEGFA*)		rs3025039	3'UTR	c.*237C>T	VEGFA-30

Table 1: Genes and their SNPs that have been found to be associated with acute mountain sickness. Also listed is the gene chromosome location, SNP location in the gene and the resulting genetic mutation as well as the linkage disequilibrium (LD) identity using in the analysis.

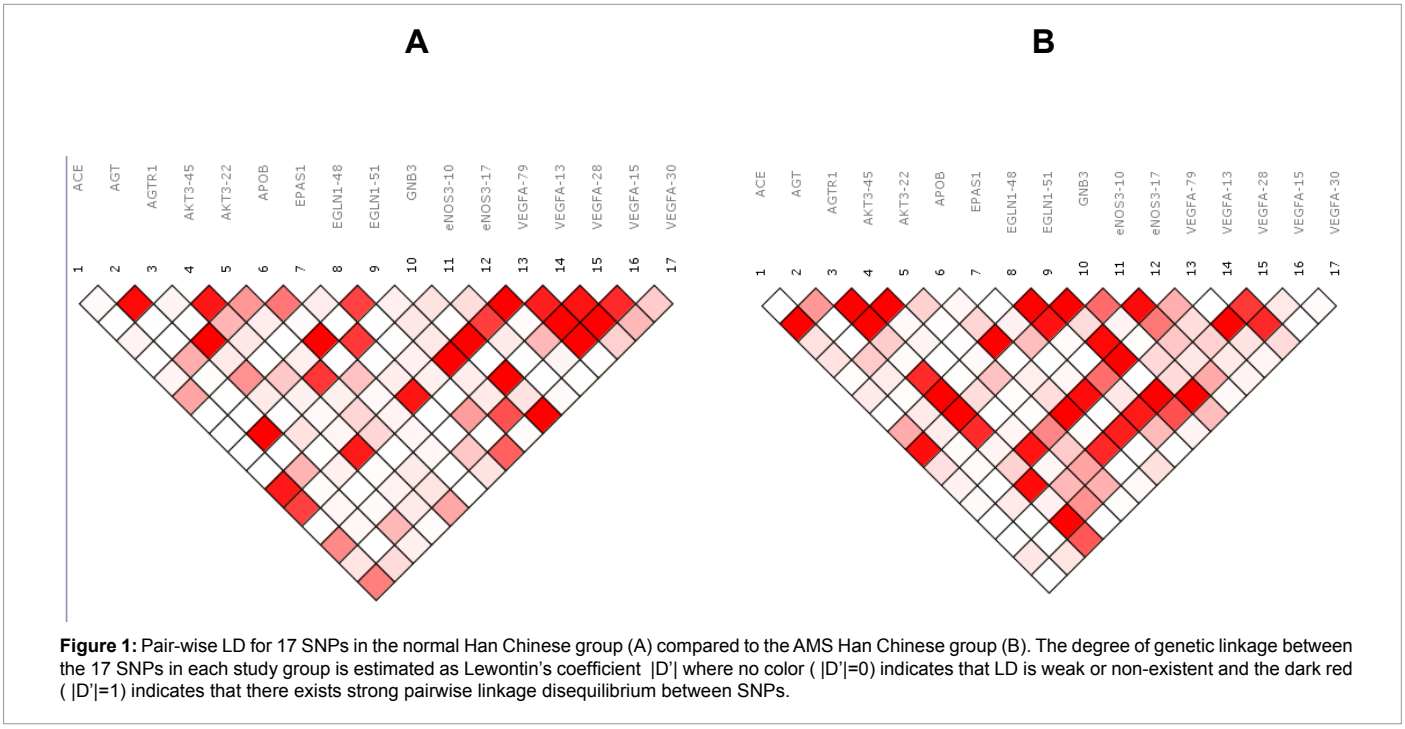

Figure 1: Pair-wise LD for 17 SNPs in the normal Han Chinese group (A) compared to the AMS Han Chinese group (B). The degree of genetic linkage between the 17 SNPs in each study group is estimated as Lewontin's coefficient |D'| where no color (|D'|=0) indicates that LD is weak or non-existent and the dark red (|D'|=1) indicates that there exists strong pairwise linkage disequilibrium between SNPs.

in genomic sequences. The TFBS and SNP location within the binding site are listed in (Table 3). TFBS that are in **bold** lettering are unique to the given allele while those with normal lettering occur with both SNP alleles. The minor allele frequencies (MAF) for the 17 SNPs were compiled from our previous studies [8-10,32] (Table 3).

Results

SNPs and punitive TFBS in the genes from renin-angiotension system (*ACE*, *AGT*, *AGTR1*), G protein signaling (*GNB3*) and LDL apolipoprotein B (*APOB*) were evaluated as well as the HIF oxygen signaling pathway (*EPAS1*&*EGLN1*) and the VEGF signaling pathway (*AKT3*, *eNOS3*&*VEGFA*) genes (Tables 1,3), (Figures 1,2). In pair-wise LD estimates the *ACE* with*AGTR1*, *EGLN1* and*GNB3* SNP pairs exhibit an increase in LD for the AMS study group compared to the control group and a decrease in *ACE* with *EPAS1*, *eNOS*-17, *VEGFA*-79, *VEGFA*-28 and *VEGFA*-30 pairs (Figure 1 and Table 2). In pair-wise LD estimates the *AGTR1* with *AKT3*-45, *AKT3*-22, *EGLN1*-48, *EGLN1*-51, *GNB3*, *eNOS3*-10, *VEGFA*-13 and *VEGFA*-15 pairs exhibit an increase in LD for the AMS study group compared to the control group and a decrease in *AGTR1* with only *APOB* (Figure 1 and Table 2). Similar results can be seen for the other pair-wise LD estimates for all the SNPs (Figure 1 and Table 2) including the association of SNPs between these signaling pathways and systems (Figure 2). The relationship between LD and SNPs can result from changes in alleles within punitive TFBS caused by the SNPs (Table 3) which sometimes results in different TFBS for each SNP allele. For example, the SNP (rs480902) *EGLN1*-T allele

S No.	Increase in LD	S No.	Decrease in LD
1	ACE vs AGTR1	1	ACE vs EPAS1
2	ACE vs EGLN1-51	2	ACE vs eNOS3-17
3	ACE vs GNB3	3	ACE vs VEGFA-79
4	AGTR1 vs AKT3-45	4	ACE vs VEGFA-28
5	AGTR1 vs AKT3-22	5	ACE vs VEGFA-30
6	AGTR1 vs EGLN1-48	6	AGT vs AGTR1
7	AGTR1 vs EGLN1-51	7	AGT vs GNB3
8	AGTR1 vs GNB3	8	AGTR1 vs APOB
9	AGTR1 vs eNOS3-10	9	AKT3-22 vs GNB3
10	AGTR1 vs VEGFA-13	10	AKT3-22 vs VEGFA-30
11	AGTR1 vs VEGFA-15	11	APOB vs EPAS1
12	AGTR1 vs VEGFA-30	12	APOB vs eNOS3-10
13	AKT3-45 vs VEGFA-15	13	EPAS1 vs GNB3
14	AKT3-45 vs VEGFA-30	14	EPAS1 vs VEGFA-15
15	AKT3-22 vs VEGFA-79	15	GNB3 vs VEGFA-79
16	AKT3-22 vs VEGFA-13	16	GNB3 vs VEGFA-30
17	AKT3-22 vs VEGFA-28	17	eNOS3-10 vs VEGFA-79
18	AKT3-22 vs VEGFA-15	18	eNOS3-17 vs VEGFA-79
19	APOB vs VEGFA-79	19	VEGFA-79 vs VEGFA-13
20	APOB vs VEGFA-28	20	VEGFA-28 vs VEGFA-15
21	EPAS1 vs VEGFA-28	21	VEGFA-28 vs VEGFA-30
22	EGLN1-48 vs GNB3	22	VEGFA-15 vs VEGFA-30
23	EGLN1-48 vs VEGFA-79		
24	EGLN1-48 vs VEGFA-28		
25	EGLN1-48 vs VEGFA-15		
26	EGLN1-51 vs GNB3		
27	EGLN1-51 vs eNOS3-17		
28	EGLN1-51 vs VEGFA-28		
29	GNB3 vs eNOS3-10		
30	eNOS3-10 vs eNOS3-17		
31	eNOS3-10 vs VEGFA-13		
32	eNOS3-10 vs VEGFA-15		

Table 2. Changes in pair-wise LD from Figure1 between 17 SNPs in the AMS Han Chinese group compared to the normal Han Chinese group.
Listed is the increase and decrease in pair-wise LD in the AMS group compared to the Han Chinese control group.

generates two unique (FOXL1 and NFE2L1::MafG) TFBS (Supplement) while the remaining potential TFBS are generated with both the *EGLN1* T and C alleles (Table 3). These changes may result in unfavorable allele or punitive TFBS combinations within haplotypes or between genes which in turn may result in disease or sickness [36]. Taken together, the changes in LD observed between these pair-wise SNP studies indicates the strong interaction of some of these SNPs with each other as a result of diversity of punitive TFBS generated by the SNP alleles (Table 3).

The *ACE*-I/D is probably the most extreme example of changes in potential TFBS and disease (Table 3). The *ACE*-D allele generates four TFBS of which only the MYB site is unique while the *ACE*-I allele generates 84 punitive TFBS with many unique TFBS including one for hypoxia TF binding (HIF1A::ARNT) and others for steroid hormone binding (PPAR, RAR, ROR and THR) (Table 3, Supplement). The increase in pair-wise LD between the *ACE* and *AGTR1* SNPs for the AMS group compared to the control group (Figure 1, Table 2) could result from the generation of the potential GATA3 TFBS created by the minor *AGTR1-C* allele since GATA3 does not occur with either of the *ACE* I/D SNP alleles (Table 3). The increase in pair-wise LD between *ACE* and *EGLN1*-51 SNPs for the AMS group compared to the normal control group could result from the potential HNF4A, HNF4G, NR2C2, NG2F1 and TAL1:GATA1 TFBS created by the common EGLN1-51 allele since these TFBS are not found to be associated with either of the *ACE* I/D SNP alleles. The increase in pair-wise LD between *ACE* and *GNB3* SNPs for the AMS group compared to the normal control group could result from the potential INSM1, NFYB and RREB1 TFBS created by the minor GNB3 allele since these TFBS are not found to be associated with the *ACE*- I/D SNP alleles. Since the *AGTR1* SNP generates only two unique potential (HOXA5 and GATA3) TFBS between the common and minor alleles, respectively (Table 3), could be the reason that the *AGTR1* SNP exhibits an increase in pair-wise LD with nine other SNPs (*AKT3*-45, *AKT3*-22, *EGLN1*-48, *EGLN*-51, *GNB3*, *eNOS3*-10, *VEGFA*-13, *VEGFA*-15, and *VEGFA*-30) whose combined alleles generate many more potential TFBS (Table 3). The SNPs of the *AKT3*, *eNOS3* and *VEGFA* genes in the VEGF signally pathway (a growth factor activator for angiogenesis) and the SNPs of the *EGLN1* and *EPAS1* genes in the HIF oxygen signaling pathway exhibit an increase in pair-wise LD for the AMS group compared to the control Han Chinese groups (Figure 1, Table 2) both within and between pathways (Figure 2). For the twelve SNPs in these two pathways there are unique potential TFBS generated between alleles in each SNP and also unique potential TFBS between all twelve SNPs (Table 3). Beside the SNP (rs4340) *ACE*-I allele, there are only two other SNPs (rs4590656 and rs3025039) *AKT3*-C and *VEGFA*-C alleles, respectively, that generate a potential unique hypoxia TF binding (HIF1A::ARNT) sites (Table 3).

A decrease in LD between SNPs for the AMS group compared to the Han Chinese normal group could indicate that some genes are responding to different attributes of the AMS sickness. As an example, the *VEGFA* SNPs (rs79469752, rs13207351, rs28357093 and rs1570360) are tightly linked within about 50bp of each other in the promoter. Consequently, they should be acting as a single linkage group as indicated in Figure 1A with the normal control group; however, in the AMS group the *VEGFA* SNPs (rs13207351 and rs1570360) are acting as though they are totally unlinked as indicated in Figure 2B. We have previously shown that these two *VEGFA* SNPs are in strong LD [9]. From an examination of Table 3, it can be seen that these two SNPs generate potential TFBS for TFs that are involved in the machinery of gene regulation (i.e. HINFP, NFIC, KLF4, KLF5, NRF1, PAX5, SP2, SPIB and ZNF354C) with no overlap in potential TFBS between the two SNPs. The decrease in LD could result from unfavorable TFBS combinations created by the alternate alleles of the *VEGFA* SNPs (rs132077351 and rs1570360) resulting in a break down in the gene regulation machinery function. An increase in LD was found in 32 pair-wise comparisons while a decrease was found 22 pair-wise comparisons between the AMS and control groups (Figure 1, Table 2).

Discussion

GWAS over the last decade have identified nearly 6,500 disease or trait-predisposing SNPs where only 7% of these are located in protein-coding regions of the genome [36,37] and the remaining 93% are located within non-coding areas [38,39] 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 [40-42]. 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 other cases, a SNP may eliminate the natural binding

Gene	ACE	AGT	AGTR1	AKT3	AKT3	APOB	EGLN1	EGLN1
SNP	rs4340	rs699	rs5186	rs4590656	rs2291409	rs693	rs480902	rs516651
Location	intron	exon	3'UTR	intron	intron	exon	intron	intron
Common allele	**I**	**T**	**A**	**C**	**G**	**C**	**T**	**C**
	FOXC1(-)	**EBF1(+)**	**HOXA5(-)**	FOXA1(+)	FOXA1(-)	**HLTF(+)**	BATF::JUN(-)	**ESRRA(-)**
	FOXC1(+)	**EBF1(-)**		**FOXD1(-)**	FOXI1(-)	**STAT4(+)**	FOS(-)	**HNF4A(+)**
	FOXL1(-)	**MZF1_5-13 (+)**		**HIF1A::ARNT(+)**	NFIC(-)	**STAT5a::STAT5b(-)**	FOS(+)	HNF4G(-)
	GATA2(+)			HNF4A(-)	**RORA_2(+)**		FOSL1(-)	**NR2C2(-)**
	HIF1A::ARNT(-)			HNF4A(+)			FOSL1(+)	**NR2F1(+)**
	PPAR(-)						FOSL2(-)	**NR4A2(-)**
	RAR(+)						FOSL2(+)	PPARG::RXRA(-)
	RORA_1(-)						**FOXL1(+)**	**RORA_1(-)**
	RORA_2(-)						JUN(var.2)(-)	**TAL1::GATA1(+)**
	THAP1(+)						JUN(var.2)(+)	USF1(+)
	THR(-)						JUNB(-)	USF2(+)
	*						JUNB(+)	
							JUND(-)	
							JUND(+)	
							NFE2L1::MafG(-)	
Minor allele	**D**	**C**	**C**	**T**	**A**	**T**	**C**	**T**
AMS (MAF)	0.42	0.45	0.05	0.401	0.5	0.085	0.547	0.071
Han (MAF)	0.29	0.33	0.02	0.383	0.424	0.055	0.437	0.055
	FOXC1(-)	**E2F1(+)**	**GATA3(-)**	FOXA1(+)	FOXA1(-)	**PAX2(-)**	BATF::JUN(-)	HNF4G(-)
	FOXL1(-)	**E2F4(+)**		HNF4A(+)	**FOXC1(+)**		FOS(-)	NFE2L1:MafG(+)
	MYB(+)	**E2F6(+)**		**HNF4G(-)**	FOXI1(-)		FOS(+)	PPARG::RXRA(-)
	THAP1(+)	**INSM1(-)**		**HOXA5(+)**	NFIC(-)		FOSL1(-)	USF1(+)
		NFE2L1::MafG(+)		**RUNX2(+)**	**NR4A2(+)**		FOSL1(+)	USF2(+)
				SOX10(+)			FOSL2(-)	
				SPIB(-)			FOSL2(+)	
				TCF7L2(-)			JUN(var.2)(-)	
							JUN(var.2)(+)	
							JUNB(-)	
							JUNB(+)	
							JUND(-)	
							JUND(+)	

EPAS1	GNB3	eNOS3	eNOS3	VEGFA	VEGFA	VEGFA	VEGFA	VEGFA
Unknown	rs2071057	rs1007311	rs1799983	rs79469752	rs13207351	rs28357093	rs1570360	rs3025039
intron	5'UTR	intron	exon	promoter	promoter	promoter	promoter	3'UTR
C	**G**	**A**	**G**	**C**	**G**	**A**	**G**	**C**
BATF::JUN(-)	E2F4(-)	**BRCA1(+)**	EBF1(+)	**E2F1(+)**	**NFIC(+)**	**EBF1(+)**	**EGR1(-)**	BATF::JUN(-)
JUN(var.2)(-)		INSM1(-)	**FOXL2(-)**	**NFIC(-)**	NRF1(-)	**E2F3(-)**	KLF5(-)	**BRCA1(+)**
JUN::FOX(-)		**MZF1_1-4(-)**	**GATA3(+)**	PAX5(-)		NRF1(+)	**MZF1_5-13(+)**	E2F6(-)
JUN::FOX(+)		RFX5(+)	INSM1(-)	**TFAP2C(+)**			SP1(-)	**ESR2(+)**
RUNX1(+)		RUNX1(-)	NFE2L1::MafG(+)				**SP2(-)**	**HIF1A::ARNT(-)**
SOX10(+)		TFAP2A(+)	**RUNX1(-)**				SREBF1(-)	**NFE2L1::MafG(+)**
		TFAP2C(+)	TAL1::GATA1(-)				TFAP2C(-)	
			THAP1(+)				ZNF354C(-)	
			ZEB1(-)					
G	**A**	**G**	**T**	**T**	**A**	**C**	**A**	**T**
0.218	0.05	0.418	0.228	0.077	0.313	0.16	0.196	0.155
0.26	0	0.463	0.102	0.03	0.33	0.11	0.2	0.125
BATF::JUN(-)	**E2F1(-)**	**EBF1(-)**	EBF1(-)	**ESR2(+)**	**HINFP(+)**	NRF1(-)	**EGR2(-)**	BATF::JUN(-)
ELK1(-)	E2F4(-)	**EBF1(+)**	EBF1(+)	**NFE2L1::MafG(-)**	NRF1(-)	NRF1(+)	**EHF(-)**	E2F6(-)
JUN(var.2)(-)	**E2F6(-)**	INSM1(-)	INSM1(-)	**NR2C2(+)**	NRF1(+)	**RFX5(+)**	**FOXH1(-)**	**NFE2::MAF(-)**
JUN::FOX(-)	**INSM1(-)**	**MZF1_5-13(-)**	**MZF1_1-4(-)**	**NR2F1(-)**	**PAX5(-)**	**THAP1(+)**	KLF5(-)	**RFX5(+)**

NRF1(+)	**MZF1_5-13(-)**	**NFE2L1::MafG(+)**	**MZF1_5-13(-)**	PAX5(-)			**MAFK(-)**	**YY1(+)**
SOX10(+)	**NFIC(-)**	RFX5(+)	NFE2L1::MafG(+)	**RUNX1(+)**			SP1(-)	
TP53(-)	**NFYB(+)**	RUNX1(-)	TAL1::GATA1(-)	**THAP1(-)**			**SPIB(+)**	
	RREB1(+)	TFAP2A(+)	**TAL1::TCF3(-)**				SREBF1(-)	
	THAP1(+)	TFAP2C(+)	**USF1(+)**				TFAP2C(-)	
			USF2(+0				**THAP1(-)**	
			ZNF263(-)				ZNF354C(-)	
			ZNF354C(+)					

Table 3. Location of gene SNPs contained in potential TFBS. (+/-) is the DNA strand location of the TFBS.
TFBS in **bold** are only present for the given allele.
*A total of 84 punitive TFBS for the ACE-I allele (see Appendix 1). MAF is the minor allele frequency.

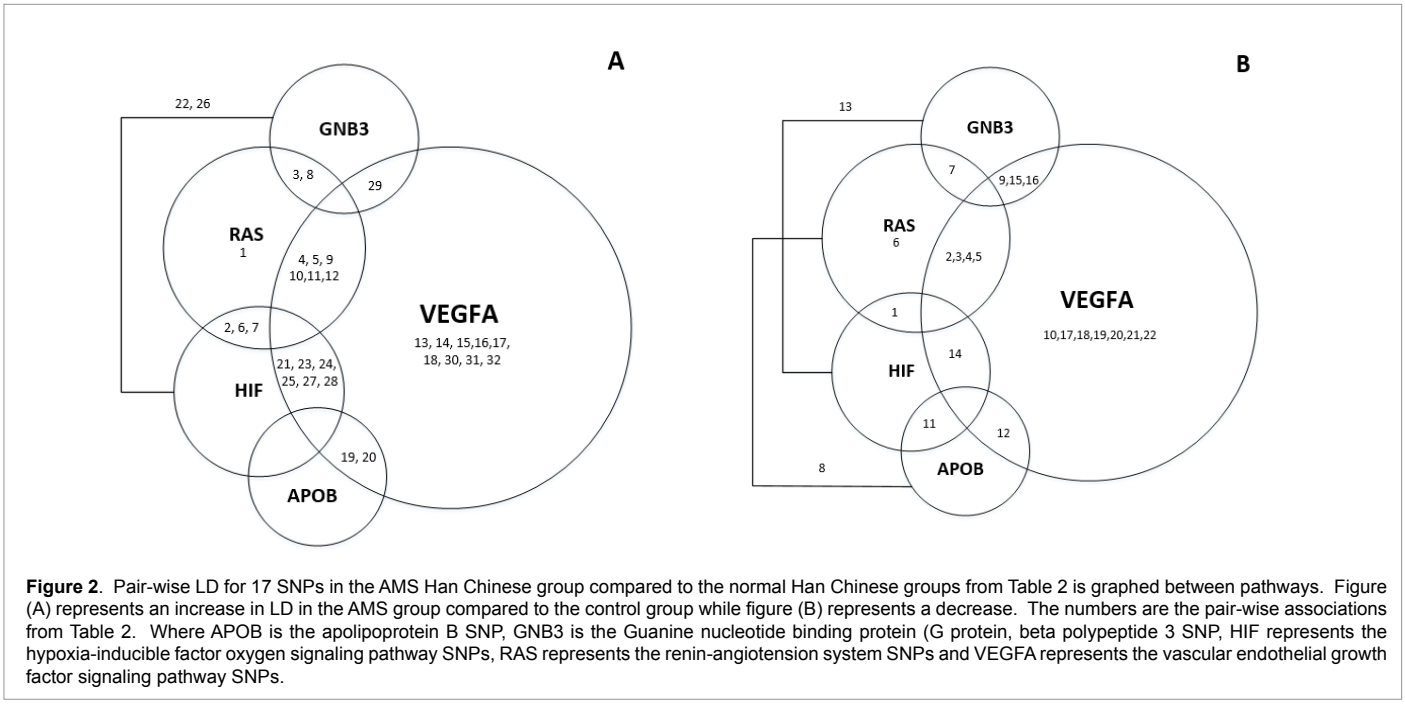

Figure 2. Pair-wise LD for 17 SNPs in the AMS Han Chinese group compared to the normal Han Chinese groups from Table 2 is graphed between pathways. Figure (A) represents an increase in LD in the AMS group compared to the control group while figure (B) represents a decrease. The numbers are the pair-wise associations from Table 2. Where APOB is the apolipoprotein B SNP, GNB3 is the Guanine nucleotide binding protein (G protein, beta polypeptide 3 SNP, HIF represents the hypoxia-inducible factor oxygen signaling pathway SNPs, RAS represents the renin-angiotension system SNPs and VEGFA represents the vascular endothelial growth factor signaling pathway SNPs.

site or generate a new binding site. In which cases the gene is no longer regulated by the original TF. Therefore, functional regulatory(r) SNPs in TFBS may result in differences in gene expression, phenotypes and susceptibility to environmental exposure [42]. Examples of rSNPs associated with disease susceptibility are numerous and several reviews have been published [42-45]. rSNPs which occur in the non-coding regions of these genes have been found to be associated with human diseases or sicknesses. These non-coding regions host the binding sites for the transcription factors that regulate gene expression [36].

We have previously reported on SNPs, potential TFBS and high altitude sickness (HAS) [31]. In this report we include the association of SNPs and LD with potential TFBS inAMS. AMS occurs in up to 50% of individuals ascending to high altitude[46] and may progress to life-threatening pulmonary and cerebral edema in a minority of cases [47]. One of the most well studied gene polymorphisms in association with HAS is the angiotensin-converting enzyme (*ACE*) insertion/deletion (rs4340) from the RAS system [48-50]. The *ACE*-I allele has been associated with superior performance benefit for mountaineers ascending to extreme altitude compared to the *ACE*-D allele [51,52], while the *ACE* -I/I genotype has been found to maintain higher arterial oxygen saturation at rest and during exercise at high altitude [53]. Perhaps the genetic reason for these findings is that the 288bp *ACE*-I allele generates at least 84 potential TFBS compared to the *ACE*-D allele which generates four TFBS (Table 3). Of these 84 punitive TFBS for the *ACE*-I allele, there is a hypoxia TF binding (HIF1A::ARNT) site and other TFBS for steroid hormone binding proteins (PPAR, RAR, ROR and THR). From a scan of the entire ACE gene with the VNTI program we find that the HIF1A::ARNT binding site occurs only once in the ACE-I allele and the ROR site occurs one other time in intron 14, while the PPAR, RAR and THR binding sites occur multiple times. The HIF1A::ARNT binding site would be beneficial to enhanced arterial oxygen saturation of red blood cells and superior performance in exercise at high altitudes. From our previous study, we found that there is a much higher incidence of *ACE*-D allele (0.42) in AMS patients than the normal Han Chinese control roup (0.29) [8] (Table 3) suggesting that the *ACE*-I allele does better in high altitude environments than the *ACE*-D allele.Also in the RAS system, the angiotensinogen (*AGT*) M268T polymorphism (rs699) has been reported to be significantly associated with HAPE in a Chinese population [54] and more recently with an Indian population [55]. The *AGT*-T allele (M268) generates two unique potential TFBS (EBF1 and MZF1_5-13) while the *AGT*-C allele (268T) generates five unique potential TFBS (E2F1, E2F4, E2F6, INSM1 and NFE2L1::MafG) (Table 3). Only the dimer MFE2L1::MafG TFs are involved with the activation of α and β globin and erythrocyte development (Supplement) which would be important in high altitude (HA) environments and consequently should benefit people carrying

the *AGT*-C allele (268T). Another gene in the RAS system is the AGTR1 whose polymorphism A1166C (rs5186) has been associated with many human diseases [56-59]. The AGTR1-A allele generates one unique potential TFBS for the HOXA5 TF while the AGTR1-C allele generates a different unique potential TBFS for the GATA3 TF (Table 3). HOXA5 is part of a developmental regulatory system and GATA3 plays an important role in endothelial cell biology (Supplement).

Although the *APOB* polymorphism (rs693) has not been linked to HAS, it been found to be associated with other human diseases such as dyslipidemia and higher LDL levels [60] and has been shown to influence plasma levels [61, 62]. The *APOB*-C allele generates three unique potential TFBS for the HLTF, STAT4 and STAT5a::STAT5b TFs while the APOB-T allele generates one unique potential TFBS for the PAX2 TF (Supplement). For the *GNB3* polymorphism (rs2071057), we have reported a significantly high incidence of the G-protein (GNB3) (-350)A allele in the AMS patients [8]. The *GNB3*-G allele generates no unique potential TFBS but the GNB3-A allele generates eight unique TFBS for the E2F1, E2F6, INSM1, MZF1_5-13, NFIC, NFYB, RREB1 and THAP1 TFs. The RREB1 TF is involved with repressing the angiotensinogen gene which regulates blood pressure and fluid balance (Supplement). In the study, we found no *GNB3*-A alleles in the control Han Chinese study group but did find a 5% occurrence of this allele in the AMS Han Chinese group [8].

The *EPAS1* gene from the HIF oxygen signaling pathway has a SNP whose *EPAS1*-C allele generates one unique potential TFBS for the RUNX1 TF while the *EPAS1*-T allele generates three unique potential TFBS for the ELK1, NRF1 and TP53 TFs. The NRF1 TF actives nuclear genes required for respiration, heme biosynthesis, and mitochondrial DNA transcription and replication while the TP53 TF is involved with response to cellular stress (such as DNA damage, hypoxia, spindle damage). Both of which would affect individuals in HA environments. The *EGLN1* gene is also part of the HIF oxygen signaling pathway. The *EGLN1*-T allele of the rs480902 polymorphism generates two unique potential TFBS for the FOXL1 and NFE2L1: MafG TFs while the *EGLN1*-C allele does not generate any unique TFBS; however, the two rs480902 SNP alleles generate seven common TFBS (Table 3). The FOXL1 TF is a target gene of the hedgehog signaling pathway (Supplement) which is a key regulatory of vertebrate organogenesis and the TF is involved with proper proliferation and differentiation in the gastrointestinal epithelium. The dimer MFE2L1::MafG TFs is involved with the activation of α and β globin and erythrocyte development (Supplement) which would be important in HA environments and consequently should benefit people carrying the *EGLN1*-T allele.The *EGLN1*-C allele of the rs516651 polymorphism generates seven unique potential TFBS for the ESRRA, HNF4A, NR2C2, NR2F1, NR4A2, RORA_1 and TAL1::GATA1 TFs while the *EGLN1*-T allele does not generate any unique TFBS; however, the two rs516651 SNP alleles generate five common TFBS (Table 3). Of the seven unique TFBS for the *EGLN1*-C allele, the ESRRA, NR2C2, NR2F1 and RORA_1 TFs are involved with steroid hormone activity including estrogen receptor (ER), PPAR and THR at this binding site (Supplement). The interaction between EGLN1 and PPAR has been well documented in HA environments [21,63-65].

The *AKT3* gene from the VEGF signaling pathway has two SNPs (rs4590656 and rs2291409) whose alleles generate unique punitive TFBS. The *AKT3*-C allele from the rs4590656 polymorphism generates two unique potential TBFS for the FOXD1 and HIF1A::ARNT TFs while the *AKT3*-T allele generates six unique potential TFBS for the HNF4G, HOXA5, RUNX2, SOX10, SPIB and TCF7L2 TFs.The FOXD1 TF is involved with positional identity in the developing retina while the dimer HIF1A::ARNT TF plays an essential role in cellular and systemic responses to hypoxia (Supplement). The HNF4G TF is involved in steroid hormone receptor activity and sequence-specific DNA binding transcription factor activity while the TCF7L2 TF has been implicated in blood glucose homeostasis. The *AKT3*-G allele from the rs2291409 polymorphism generates one unique potential TFBS for the RORA_2 TF which regulates a number of genes involved with lipid metabolism, in cerebellum and photoreceptor development and skeletal muscle development. The *AKT3*-A allele from this polymorphism generates two unique potential TFBS for the FOXC1 and NR1A2 TFs. While the FOXC1 TF is an important regulator of cell viability and resistance to oxidative stress in the eye, the NR4A2 TF is a regulatory for differentiation and maintenance of meso-diencephalic dopaminergic neurons during development (Supplement).

The *eNOS3* gene also from the VEGF signaling pathway has two SNPs (rs1007311 and rs1799983) whose alleles generate unique TFBS. The *eNOS3*-A allele from the rs1007311 polymorphism generates two unique potential TFBS for the BRCA1 and MZF1_1-4 TFs while the eNOS3-G allele generates three unique potential TFBS for the EBF1, MZF1_5-13 and NFE2L1::MafG TFs. The two alleles of the rs1007311 SNP also generate five common potential TFBS for the INSM1, RFX5, RUNX1, TFAP2A and TFAP2C TFs (Table 3). The BRCA1 TF plays a role in maintaining genomic stability while the MZF1_1-4 TF functions as a transcription regulator (Supplement). The EBF1 TF is involved with transcription machinery while the MZF1_5-13 TF is a regulator of transcriptional events during hemopoietic development and the NFE2L1::MafG TF coordinates the up-regulation of cytoprotective genes. The *eNOS3*-G allele from the rs1799983 polymorphism generates five unique potential TFBS for the FOXL2, GATA3, RUNX1, THAP1 and ZEB1 TFs while the eNOS3-T allele generates seven unique potential TFBS for the MZF1_1-4, MZF1_5-13, TAL1::TCF3, USF1 & 2, ZNF263 and ZNF354C TFs (Table 3). The two alleles of the rs1799983 polymorphism generate four common potential TFBS for the EBF1, INSM1, NFE2L1::MafG and TAL1::GATA1 TFs. Perhaps the TF with the most effect on AMS would be RUNX1 which is involved in the development of normal hematopoiesis. The potential TFBS (TTCTGGGGGCT) for RUNX1 is generated by the rs1799983 e*NOS3*-G allele whose frequency is 0.772 in the AMS group compared to 0.898 in the normal Han control group (Table 3).Since the RUNX1 binding site (TTCTGGGGCTG) commonly occurs among Han Chinese with the rs1007311 polymorphism, might explain why an increase in LD is seen between the two eNOS3 SNPs (Figure 1, Table 2). The potential RUNX1 TFBS (TTCTGGGGCTG) at the rs1007311 polymorphism is the only motifthat occurs in the gene from a scan of eNOS3 with the Vector NTI Advance 11 computer program.

The five *VEGFA* SNPs included in this study are rs79469742, rs13207351, rs28357093, rs1570360 and rs3025039 whose allele frequencies and unique potential TFBS for each SNP are found in Table 3. The *VEGFA*-C allele from the rs79469742 polymorphism generates three unique potential TFBS for the E2F1, NFIC and TFAP2C TFs while the *VEGFA*-T allele generates six unique potential TFBS for the ESR2, NFE2L1::MafG, NR2C2, NR2F1, RUNX1 and THAP1 TFs. The two alleles of this SNP generate one common potential TFBS for the PAX5 TF. Perhaps the most interesting unique potential TFBS generated by the *VEGFA*-T allele are for the ESR2 and NR2C2 TFs which involve the estrogen nuclear receptor and regulation of the nuclear receptor signaling pathways for steroid hormone. Another unique potential TFBS created by this allele is for the RUNX1 TF which is involved in the development of normal hematopoiesis (Table 3).The *VEGFA*-G

allele from the rs13207351 polymorphism generates one unique potential TFBS for the NFIC TF while the *VEGFA*-A allele generates two unique potential TFBS for the HINFP and PAX5 TFs. These TFs are involved with transcription machinery (Table 3, Supplement). The two alleles of this SNP generate one common potential TFBS for the NRF1 TF which is involved with respiration, heme biosynthesis, and mitochondrial DNA transcription and replication. The *VEGFA*-A allele of the rs28357093 polymorphism creates two unique potential TFBS for the EBF1 and E2F3 TSs while the *VEGFA*-C allele generates two unique potential TFBS for the RFX5 and THAP1 TFs. These TFs are also involved with transcription machinery (Table 3, Supplement). The two alleles of this SNP also generate one common potential TFBS for the NRF1 TF those function is mentioned above.

The *VEGFA*-G allele of the rs1570360 polymorphism generates four unique potential TFBS for the EGR1, KLF4, MZF1_5-13 and SP2 TFs while the *VEGFA*-A allele generates six unique potential TFBS for the EGR2, EHF, FOXH1, MAFK, SPIBand THAP1TFs. The two alleles of this SNP also generate five common potential TFBS for the KLF5, SP1, SREBF1, TFAP2C and ZNF354C TFs. The potential TFBS created by this SNP are all involved with transcriptional regulation (Table 3, Supplement).The *VEGFA*-C allele of the rs3025039 polymorphism generates four unique potential TFBS for the BRCA1, ESR2, HIF1A::ARNT and NFE2L1::MafG TFs while the *VEGFA*-T allele generates three unique potential TFBS for the NFE2::MAF, RFX5 and YY1 TFs. The two alleles of this SNP also generate two common potential TFBS for the BATF::JUN and E2F6 TFs. Of all these potential TFBS, perhaps the TF with the most effect on AMS would be HIF1A::ARNT which plays an essential role in cellular and systemic responses to hypoxia which is generated by the *VEGFA*-C allele whose frequency is 0.845 in the AMS group compared to 0.875 in the normal Han control group (Table 3).

In conclusion, since nearly all of the SNPs used in this study have previously been found to be associated with AMS, it may not be only one SNP that alters the TFBS for a TF to bind that causes the sickness but more likely a combination of SNP changes in TFBS that lead to the illness. Perhaps, the *ACE* (I/D) rs4340 SNP would be the largest contributor for AMS because the *ACE*-I allele creates 84 punitive TFBS compared to the *ACE*-D allele that creates only four TFBS. The interaction of all SNPs from different areas of the genome (Table 1) as examined by LD analysis (Figure 1, Table 2) indicates that certain TFBS associations throughout the genomeare involved in AMS (Table 2). SNPs that alter the TFBS are not only found in the promoter regions but in the introns, exons and the UTRs of a gene (Table 3). The nucleus of the cell is where epigenetic alterations and TFs operate to convert chromosomes into single stranded DNA for mRNA transcription while it is the cytoplasm where mRNA is processed by separating exons and introns for protein translation. Consequently, it doesn't matter where TFs bind the DNA in the nucleus because it is only there that TFs function. The SNPs outlined in this report should be considered as rSNPs since they change the DNA landscape for TF binding and have been associated with AMS.

Dedication

This manuscript is dedicated to the work and memory of Xue-Han Ning.

References

1. Hackett PH, Roach RC (2001) High-altitude illness. N Engl J Med 345: 107-114.
2. Bartsch P, Bailey DM, Berger MM, Knauth M, Baumgartner RW (2004) Acute mountain sickness: controversies and advances. High Alt Med Biol 5: 110-124.
3. Ning XH, Li SP (2006) Plateau Tin Road Health Line (Universal Self-Care Reader). Shanghai Science and Technology Publishing House, Shanghai, China, 66-8.
4. West JB; American College of Physicians; American Physiological Society (2004) The physiologic basis of high-altitude diseases. Ann Intern Med 141: 789-800.
5. Schoene RB (2008) Illnesses at high altitude. Chest 134: 402-416.
6. Strohl KP (2008) Lessons in hypoxic adaptation from high-altitude populations. Sleep Breath 12: 115-121.
7. Wilson MH, Newman S, Imray CH (2009) The cerebral effects of ascent to high altitudes. Lancet Neurol 8: 175-191.
8. Buroker NE, Ning XH, Zhou ZN, Li K, Cen WJ, et al. (2010) Genetic associations with mountain sickness in Han and Tibetan residents at the Qinghai-Tibetan Plateau. Clin Chim Acta 411: 1466-1473.
9. Buroker NE, Ning XH, Zhou ZN, Li K, Cen WJ, et al. (2012) AKT, ANGPTL, eNOS, and VEGFA associations with high altitude sickness in Han and Tibetan Chinese at the Qinghai-Tibetan Plateau. Int J Hematol 96: 200-213.
10. Buroker NE, Ning XH, Zhou ZN, Li K, Cen WJ, et al. (2012) EPAS1 and EGLN1 associations with high altitude sickness in Han and Tibetan Chinese at the Qinghai-Tibetan Plateau. Blood cells molecules & diseases 49: 67-73.
11. Buroker NE, Ning XH, Zhou ZN, Li K, Cen WJ, et al. (2013) VEGFA SNPs and transcriptional factor binding sites associated with high altitude sickness in Han and Tibetan Chinese at the Qinghai-Tibetan Plateau. J Physiol Sci 63: 183-93.
12. Naber CK, Siffert W (2004) Genetics of human arterial hypertension. Minerva Med 95: 347-356.
13. Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, et al. (2008) Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. Nat Genet 40: 189-197.
14. Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, et al. (2008) Newly identified loci that influence lipid concentrations and risk of coronary artery disease. Nat Genet 40: 161-9.
15. Hamm HE (1998) The many faces of G protein signaling. J Biol Chem 273: 669-672.
16. Siffert W, Rosskopf D, Siffert G, Busch S, Moritz A, et al. (1998) Association of a human G-protein beta3 subunit variant with hypertension. Nat Genet 18: 45-48.
17. Rosskopf D, Manthey I, Siffert W (2002) Identification and ethnic distribution of major haplotypes in the gene GNB3 encoding the G-protein beta3 subunit. Pharmacogenetics 12: 209-220.
18. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, et al. (2010) Sequencing of 50 human exomes reveals adaptation to high altitude. Science 329: 75-8.
19. Aggarwal S, Negi S, Jha P, Singh PK, Stobdan T, et al. (2010) EGLN1 involvement in high-altitude adaptation revealed through genetic analysis of extreme constitution types defined in Ayurveda. Proceedings of the National Academy of Sciences of the United States of America 107: 18961-6.
20. Bigham A, Bauchet M, Pinto D, Mao X, Akey JM, et al. (2010) Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. PLoS genetics 6.
21. Simonson TS, Yang Y, Huff CD, Yun H, Qin G, et al. (2010) Genetic evidence for high-altitude adaptation in Tibet. Science 329: 72-75.
22. Xu S, Li S, Yang Y, Tan J, Lou H, et al. (2011) A genome-wide search for signals of high-altitude adaptation in Tibetans. Mol Biol Evol 28: 1003-1011.
23. Lewontin RC, Kojima K (1960) The Evolutionary Dynamics of Complex Polymorphisms. Evolution 14: 458-72.
24. Buroker NE (2014) VEGFA rSNPs, transcriptional factor binding sites and human disease. J Physiol Sci 64: 73-76.
25. Buroker NE (2013) AKT3 rSNPs, Transcritional Factor Binding Sites and Human Disease. Open Journal of Blood Diseases 3: 116-29.
26. Buroker NE (2013) ATF3 rSNPs, transcriptional factor binding sites and human etiology. Open Journal of Genetics 3: 253-61.
27. Buroker NE (2013) ADRBK1 (GRK2) rSNPs, Transcriptional Factor Binding Sites and Cardiovascular Disease in the Black Population. Journal of Cardiovascular Disease 2.

28. Buroker NE (2014) TBXA2R rSNPs, Transcriptional Factor Binding Sites and Asthma in Asians. Open Journal of Pediatrics 4: 148-61.

29. Buroker NE (2014) ADRBD1 (GRK2), TBXA2R and VEGFA rSNPs in KLF4 and SP1 TFBS Exhibit Linkage Disequilibrium. Open Journal of Genetics 4.

30. Hackett Phao (1992) The diagnosis accord with the Lake Louise scoring system. In: JR Sutton GC, and CS Houston (eds) Hypoxia and Mountainsickness Pergamon Press, New York, 327-30.

31. Ding K, Zhou K, He F, Shen Y (2003) LDA--a java-based linkage disequilibrium analyzer. Bioinformatics 19: 2147-2148.

32. Bryne JC, Valen E, Tang MH, Marstrand T, Winther O, et al. (2008) JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. Nucleic Acids Res 36: D102-106.

33. Sandelin A, Alkema W, Engström P, Wasserman WW, Lenhard B (2004) JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res 32: D91-94.

34. Sandelin A, Wasserman WW, Lenhard B (2004) ConSite: web-based prediction of regulatory elements using cross-species comparison. Nucleic Acids Res 32: W249-252.

35. Buroker NE (2014) Regulatory SNPs and transcriptional factor binding sites in ADRBK1, AKT3, ATF3, DIO2, TBXA2R and VEGFA. Transcription 5: e964559.

36. Pennisi E (2011) The Biology of Genomes. Disease risk links to gene regulation. Science 332: 1031.

37. Kumar V, Wijmenga C, Withoff S (2012) From genome-wide association studies to disease mechanisms: celiac disease as a model for autoimmune diseases. Semin Immunopathol 34: 567-580.

38. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, et al. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A 106: 9362-9367.

39. Kumar V, Westra HJ, Karjalainen J, Zhernakova DV, Esko T, et al. (2013) Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. PLoS Genet 9: e1003201.

40. Knight JC (2003) Functional implications of genetic variation in non-coding DNA for disease susceptibility and gene regulation. Clin Sci (Lond) 104: 493-501.

41. Wang X, Tomso DJ, Liu X, Bell DA (2005) Single nucleotide polymorphism in transcriptional regulatory regions and expression of environmentally responsive genes. Toxicol Appl Pharmacol 207: 84-90.

42. Chorley BN, Wang X, Campbell MR, Pittman GS, Noureddine MA, et al. (2008) Discovery and verification of functional single nucleotide polymorphisms in regulatory genomic regions: current and developing technologies. Mutat Res 659: 147-57.

43. Prokunina L, Alarcón-Riquelme ME (2004) Regulatory SNPs in complex diseases: their identification and functional validation. Expert Rev Mol Med 6: 1-15.

44. Buckland PR (2006) The importance and identification of regulatory polymorphisms and their mechanisms of action. Biochim Biophys Acta 1762: 17-28.

45. Sadee W, Wang D, Papp AC, Pinsonneault JK, Smith RM, et al. (2011) Pharmacogenomics of the RNA world: structural RNA polymorphisms in drug therapy. Clin Pharmacol Ther 89: 355-365.

46. Maggiorini M, Bühler B, Walter M, Oelz O (1990) Prevalence of acute mountain sickness in the Swiss Alps. BMJ 301: 853-855.

47. Roach RC, Hackett PH (2001) Frontiers of hypoxia research: acute mountain sickness. J Exp Biol 204: 3161-3170.

48. Costerousse O, Allegrini J, Lopez M, Alhenc-Gelas F (1993) Angiotensin I-converting enzyme in human circulating mononuclear cells: genetic polymorphism of expression in T-lymphocytes. Biochem J 290: 33-40.

49. Danser AH, Schalekamp MA, Bax WA, van den Brink AM, Saxena PR, et al. (1995) Angiotensin-converting enzyme in the human heart. Effect of the deletion/insertion polymorphism. Circulation 92: 1387-1388.

50. Zhu X, Bouzekri N, Southam L, Cooper RS, Adeyemo A, et al. (2001) Linkage and association analysis of angiotensin I-converting enzyme (ACE)-gene polymorphisms with ACE concentration and blood pressure. Am J Hum Genet 68: 1139-1148.

51. Montgomery HE, Marshall R, Hemingway H, Myerson S, Clarkson P, et al. (1998) Human gene for physical performance. Nature 393: 221-222.

52. Tsianos G, Eleftheriou KI, Hawe E, Woolrich L, Watt M, et al. (2005) Performance at altitude and angiotensin I-converting enzyme genotype. Eur J Appl Physiol 93: 630-633.

53. Woods DR, Pollard AJ, Collier DJ, Jamshidi Y, Vassiliou V, et al. (2002) Insertion/deletion polymorphism of the angiotensin I-converting enzyme gene and arterial oxygen saturation at high altitude. Am J Respir Crit Care Med 166: 362-366.

54. Qi Y, Niu W, Zhu T, Zhou W, Qiu C (2008) Synergistic effect of the genetic plymorphisms of the renin-angiotensin-aldosterone system on high-altitude pulmonary edema: a study from Qinghai-Tibet altitude. Eur J Epidemiol 23: 143-152.

55. Stobdan T, Ali Z, Khan AP, Nejatizadeh A, Ram R, et al. (2011) Polymorphisms of renin--angiotensin system genes as a risk factor for high-altitude pulmonary oedema. J Renin Angiotensin Aldosterone Syst 12: 93-101.

56. Liu Y, Zhuoma C, Shan G, Cui C, Hou S, et al. (2002) A1166C polymorphism of the angiotensin II type 1 receptor gene and essential hypertension in Han, Tibetan and Yi populations. Hypertens Res 25: 515-21.

57. Jones A, Dhamrait SS, Payne JR, Hawe E, Li P, et al. (2003) Genetic variants of angiotensin II receptors and cardiovascular risk in hypertension. Hypertension 42: 500-506.

58. Yan C, Zhan J, Feng W (2005) Gene polymorphisms of angiotensin II type 1 receptor and angiotensin-converting enzyme in two ethnic groups living in Zhejiang Province, China. J Renin Angiotensin Aldosterone Syst 6: 132-137.

59. Wu CK, Tsai CT, Chang YC, Luo JL, Wang YC, et al. (2009) Genetic polymorphisms of the angiotensin II type 1 receptor gene and diastolic heart failure. Journal of hypertension 27: 502-7.

60. Rodrigues AC, Sobrino B, Genvigir FD, Willrich MA, Arazi SS, et al. (2013) Genetic variants in genes related to lipid metabolism and atherosclerosis, dyslipidemia and atorvastatin response. Clin Chim Acta 417: 8-11.

61. Benn M, Stene MC, Nordestgaard BG, Jensen GB, Steffensen R, et al. (2008) Common and rare alleles in apolipoprotein B contribute to plasma levels of low-density lipoprotein cholesterol in the general population. J Clin Endocrinol Metab 93: 1038-1045.

62. Zhu H, Tucker HM, Grear KE, Simpson JF, Manning AK, et al. (2007) A common polymorphism decreases low-density lipoprotein receptor exon 12 splicing efficiency and associates with increased cholesterol. Hum Mol Genet 16: 1765-1772.

63. Storz JF (2010) Evolution. Genes for high altitudes. Science 329: 40-41.

64. Simonson TS, McClain DA, Jorde LB, Prchal JT (2012) Genetic determinants of Tibetan high-altitude adaptation. Hum Genet 131: 527-533.

65. Ge RL, Simonson TS, Cooksey RC, Tanna U, Qin G, et al. (2012) Metabolic insight into mechanisms of high-altitude adaptation in Tibetans. Mol Genet Metab 106: 244-247.

Citation: Buroker NE, Ning XH, Li K, Zhou ZN, Cen WJ, et al, (2015) SNPs, Linkage Disequilibrium and Transcriptional Factor Binding Sites Associated with Acute Mountain Sickness among Han Chinese at the Qinghai-Tibetan Plateau J Clin Med Genom 3: 120. doi: 10.4172/2472-128X.1000120

Appendix 1. The ACE-I allele creates 84 potential TFBS.

ARID3A	E2F4	FOSL2	HLF	MEF2A	NR4A2	RFX5	SREBF1	YY1
BATF::JUN	E2F6	FOXC1	HOXA5	MEF2C	NRF1	RORA_1	SREBF2	ZEB1
BRAC1	EBF1	FOXH1	IRF1	MYC::MAX	PAX5	RORA_2	SRY	ZNF263
CDX2	EGR1	FOXI1	JUN (var.2)	MZF1_1-4	PLAG1	RUNX1	STAT2::STAT1	ZNF354C
CEBPA	EHF	FOXL1	JUN::FOS	MZF1_5-13	POU2F2	RUNX2	STAT3	
CEBPB	ELK1	FOXP1	JUNB	NFATC2	PPAR	SOX10	TFAP2C	
CREB1	ELK4	GABPA	JUND	NFE2L1::MaFg	PRDM1	SOX9	THAP1	
DUX4	ESRRA	GATA2	JUND (var.2)	NFIC	RAR	SP1	THR	
E2F1	FEV	HIF1A::ARNT	KLF5	NFKB1	REL	SP2	USF1	
E2F3	FOSL1	HINFP	MAFK	NHLH1	RELA	SPIB	USF2	

Supplement. Transcriptional factors (TF), protein name and their description or function.

TF	Protein name	TF description/function
ARID3A	AT rich interactive domain 3A (BRIGHT-like)	This gene encodes a member of the ARID (AT-rich interaction domain) family of DNA binding proteins.
BATF	Basic leucine zipper transcription factor, ATF-like	The protein encoded by this gene is a nuclear basic leucine zipper protein that belongs to the AP-1/ATF superfamily of transcription factors. The leucine zipper of this protein mediates dimerization with members of the Jun family of proteins. This protein is thought to be a negative regulator of AP-1/ATF transcriptional events.
BATF::JUN	Basic leucine zipper transcription factor, ATF-like Jun proto-oncogene	The protein encoded by this gene is a nuclear basic leucine zipper protein that belongs to the AP-1/ATF superfamily of transcription factors. The leucine zipper of this protein mediates dimerization with members of the Jun family of proteins. This protein is thought to be a negative regulator of AP-1/ATF transcriptional events.
BRCA1	Breast cancer 1, early onset	This gene encodes a nuclear phosphoprotein that plays a role in maintaining genomic stability, and it also acts as a tumor suppressor.
CDX2	Caudal type homeobox 2	This gene is a member of the caudal-related homeobox transcription factor gene family. The encoded protein is a major regulator of intestine-specific genes involved in cell growth an differentiation.
CEBPA	CCAAT/enhancer binding protein (C/EBP), alpha	C/EBP is a DNA-binding protein that recognizes two different motifs: the CCAAT homology common to many promoters and the enhanced core homology common to many enhancers
CEBPB	CCAAT/enhancer binding protein (C/EBP), beta	Important transcriptional activator regulating the expression of genes involved in immune and inflammatory responses. Regulates the transcriptional induction of peroxisome proliferator-activated receptor gamma (PPARG)
COUP (NR2F1)	chicken ovalbumin upstream promoter-transcription factor	Binds to the ovalbumin promoter and, in conjunction with another protein (S300-II) stimulates initiation of transcription. Binds to both direct repeats and palindromes of the 5'-AGGTCA-3' motif.
CREB1	cAMP responsive element binding protein 1	Phosphorylation-dependent transcription factor that stimulates transcription upon binding to the DNA cAMP response element (CRE), a sequence present in many viral and cellular promoters.
DUX4	double homeobox 4	Involved in transcriptional regulation. May regulate microRNA (miRNA) expression
E2F1-6	E2F transcription factors 1-6	The protein encoded by this gene is a member of the E2F family of transcription factors. The E2F family plays a crucial role in the control of cell cycle and action of tumor suppressor proteins and is also a target of the transforming proteins of small DNA tumor viruses. The E2F proteins contain several evolutionally conserved domains found in most members of the family. These domains include a DNA binding domain, a dimerization domain which determines interaction with the differentiation regulated transcription factor proteins (DP), a transactivation domain enriched in acidic amino acids, and a tumor suppressor protein association domain which is embedded within the transactivation domain.
EBF1	Transcription factor COE1	EBF1 has been shown to interact with ZNF423 and CREB binding proteins.
EGR1	Early growth response 1	The protein encoded by this gene belongs to the EGR family of C2H2-type zinc-finger proteins. It is a nuclear protein and functions as a transcriptional regulator. The products of target genes it activates are required for differentitation and mitogenesis.

EGR2	Early growth response 2	The protein encoded by this gene is a transcription factor with three tandem C2H2-type zinc fingers Sequence-specific DNA-binding transcription factor.
EHF	Ets homologous factor	This gene encodes a protein that belongs to an ETS transcription factor subfamily characterized by epithelial-specific expression (ESEs). The encoded protein acts as a transcriptional repressor and may be involved in epithelial differentiation and carcinogenesis.
ELF5	E74-like factor 5	A member of an epithelium-specific subclass of the Ets transcritpion factor family.
ELK1	ELK1, member of ETS oncogene family	This gene is a member of the Ets family of transcription factors and of the ternary complex factor (TCF) subfamily. The protein encoded by this gene is a nuclear target for the ras-raf-MAPK signaling cascade.
ELK4	ELK4, ETS-domain protein (SRF accessory protein 1)	This gene is a member of the Ets family of transcription factors and of the ternary complex factor (TCF) subfamily. Proteins of the TCF subfamily form a ternary complex by binding to the the serum response factor and the serum reponse element in the promoter of the c-fos proto-oncogene.
ESR2	Estrogen receptor beta	Estrogen receptor β is a member of the family of estrogen receptors and the superfamily of nuclear receptor transcription factors and is expressed by many tissues including blood monocytes and tissue macrophages, colonic and pulmonary epithelial cells.
ESRRA	Estrogen-related receptor alpha	This nuclear receptor acts as a site-specific transcription regulator and has been also shown to interact with estrogen and the transcripton factor TFIIB by direct protein-protein contact. The binding and regulatory activities of this protein have been demonstrated in the regulation of a variety of genes including lactoferrin, osteopontin, medium-chain acyl coenzyme A dehydrogenase (MCAD) and thyroid hormone receptor genes.
ETS1	Protein C-ets-1	The protein encoded by this gene belongs to the ETS family of transcription factors and has been shown to interact with TTRAP, UBE2I and Death associated protein.
FEV	ETS oncogene family	It functions as a transcriptional repressor.
FOS	FBJ murine osteosarcoma viral oncogene homolog	The **Fos** gene family consists of 4 members: **FOS**, FOSB, FOSL1, and FOSL2. These genes encode leucine zipper proteins that can dimerize with proteins of the JUN family, thereby forming the transcription factor complex AP-1. As such, the **FOS** proteins have been implicated as regulators of cell proliferation, differentiation, and transformation. In some cases, expression of the **FOS** gene has also been associated with apoptotic cell death.
FOSL1 & 2	FOS-like antigen 1 & 2	GO annotations related to this gene include *RNA polymerase II regulatory region sequence-specific DNA binding* and sequence-specific DNA binding transcription factor activity.
FOXA1	Forkhead box A1	Transcription factor that is involved in embryonic development, establishment of tissue-specific gene expression and regulation of gene expression in differentiated tissues. Is thought to act as a 'pioneer' factor opening the compacted chromatin for other proteins through interactions with nucleosomal core histones and thereby replacing linker histones at target enhancer and/or promoter sites. Involved in the development of multiple endoderm-derived organ systems such as liver, pancreas, lung and prostate. Modulates the transcriptional activity of nuclear hormone receptors.
FOXC1	Forkhead box C1	An important regulator of cell viability and resistance to oxidative stress in the eye.
FOXD1	Forkhead box D1	Transcription factor required for formation of positional identity in the developing retina, regionalization of the optic chiasm and morphogenesis of the kidney.
FOXH1	Forkhead box H1	Transcriptional activator
FOXI1	Forkhead box I1	Transcriptional activator required for the development of normal hearing, sense of balance and kidney function.
FOXL1	Forkhead box L1	Transcription factor required for proper proliferation and differentiation in the gastrointestinal epithelium. Target gene of the hedgehog (Hh) signaling pathway.
FOXL2	Forkhead box L2	Transcriptional regulator. Critical factor essential for ovary differentiation and maintenance, and repression of the genetic program for somatic testis determination.
FOXP1	Forkhead box P1	This gene belongs to subfamily P of the forkhead box (FOX) transcription factor family. Forkhead box transcription factors play important roles in the regulation of tissue- and cell type-specific gene transcription during both development and adulthood. Transcriptional repressor. It plays an important role in the specification and differentiation of lung epithelium.
GABPA	GA-binding protein alpha chain	One of three GA-binding protein transcription factor subunits which functions as a DNA-binding subunit which shares identity with a subunit encoding the nuclear respiratory factor 2 gene and is likely involved in activation of cytochrome oxidase expression and nuclear control of mitochondrial function.
GATA2	GATA binding protein 2	A member of the GATA family of zinc-finger transcription factors that are named for the consensus nucleotide sequence they bind in the promoter regions of target genes and play an essential role in regulating transcription of genes involved in the development and proliferation of hematopoietic and endocrine cell lineages.

GATA2	GATA binding protein 2	A member of the GATA family of zinc-finger transcription factors that are named for the consensus nucleotide sequence they bind in the promoter regions of target genes and play an essential role in regulating transcription of genes involved in the development and proliferation of hematopoietic and endocrine cell lineages.
GATA3	GATA binding protein 3	Plays an important role in endothelial cell biology.
HIF1A:ARNT	Hypoxia-inducible factor 1:Aryl hydrocarbon receptor nuclear translocator	HIF1 is a homodimeric basic helix-loop-helix structure composed of HIF1a, the alpha subunit, and the aryl hydrocarbon receptor nuclear translocator (Arnt), the beta subunit. The protein encoded by HIF1 is a Per-Arnt-Sim (PAS) transcription factor found in mammalian cells growing at low oxygen concentrations. It plays an essential role in cellular and systemic responses to hypoxia.
HINFP	Histone H4 transcription factor	Transcriptional repressor that binds to the consensus sequence 5'-CGGACGTT-3' and to the RB1 promoter. Transcriptional activator that promotes histone H4 gene transcription at the G1/S phase transition in conjunction with NPAT. Also activates transcription of the ATM and PRKDC genes. Autoregulates its expression by associating with its own promoter
HLF	Hepatic leukemia factor	This gene encodes a member of the proline and acidic-rich (PAR) protein family, a subset of the bZIP transcription factors. The encoded protein forms homodimers or heterodimers with other PAR family members and binds sequence-specific promoter elements to activate transcription.
HLTF	Helicase-like transcription factor	Member of the SWI/SNF (SWItch/Sucrose NonFermentable) family which have helicase and ATPase activities and are thought to regulate transcription of certain genes by altering the chromatin.
HNF4A	Hepatocyte nuclear factor 4, alpha	The protein encoded by this gene is a nuclear transcription factor which binds DNA as a homodimer. The encoded protein controls the expression of several genes, including hepatocyte nuclear factor 1 alpha, a transcription factor which regulates the expression of several hepatic genes. This gene may play a role in development of the liver, kidney, and intestines
HNF4G	Hepatocyte nuclear factor 4, gamma	Steroid hormone receptor activity and sequence-specific DNA binding transcription factor activity. An important paralog of this gene is RXRA.
HOXA5	Homeobox protein Hox-A5	DNA-binding transcription factor which may regulate gene expression, morphogenesis, and differentiation.
ISNM1	Insulinoma-associated 1	Insulinoma-associated 1 gene is intronless and encodes a protein containing both a zinc finger DNA-binding domain and a putative prohormone domain. This gene is a sensitive marker for neuroendocrine differentiation of human lung tumors.
IRF1,2	Interferon regulatory factor	Members of the interferon regulatory transcription factor (IRF) family that contain a conserved N-terminal region of about 120 amino acids, which folds into a structure that binds specifically to the interferon consensus sequence (ICS).
JUN JUN(var.2)	Jun Proto-Oncogene	This gene is the putative transforming gene of avian sarcoma virus 17. It encodes a protein which is highly similar to the viral protein, and which interacts directly with specific target DNA sequences to regulate gene expression.
JUNB	Jun B proto-oncogene	Transcription factor involved in regulating gene activity following the primary growth factor response.
JUND JUND (var.2)	Jun D proto-oncogene	The protein encoded by this intronless gene is a member of the JUN family, and a functional component of the AP1 transcription factor complex. This protein has been proposed to protect cells from p53-dependent senescence and apoptosis.
JUN::FOS	*Jun* proto-oncogene FBJ murine osteosarcoma viral oncogene homolog	Promotes activity of NR5A1 when phosphorylated by HIPK3 leading to increased steroidogenic gene expression upon cAMP signaling pathway stimulation. Has a critical function in regulating the development of cells destined to form and maintain the skeleton. It is thought to have an important role in signal transduction, cell proliferation and differentiation.
JUN::FOX	Jun B proto-oncogene Forkhead box	Heterodimer TF binding site
KLF5	Krueppel-like factor 5	This gene encodes a member of the Kruppel-like factor subfamily of zinc finger proteins. The encoded protein is a transcriptional activator that binds directly to a specific recognition motif in the promoters of target genes. This protein acts downstream of multiple different signaling pathways and is regulated by post-translational modification. It may participate in both promoting and suppressing cell proliferation. Expression of this gene may be changed in a variety of different cancers and in cardiovascular disease. Alternative splicing results in multiple transcript variants.
MAFK	v-maf avian musculoaponeurotic fibrosarcoma oncogene homolog K	Since they lack a putative transactivation domain, the small Mafs behave as transcriptional repressors when they dimerize among themselves. However, they seem to serve as transcriptional activators by dimerizing with other (usually larger) basic-zipper proteins and recruiting them to specific DNA-binding sites.
MEF2A	Myocyte enhancer factor 2A	The protein encoded by this gene is a DNA-binding transcription factor that activates many muscle-specific, growth factor-induced, and stress-induced genes. Mediates cellular functions not only in skeletal and cardiac muscle development, but also in neuronal differentiation and survival.
MEF2C	Myocyte enhancer factor 2C	Transcription activator which binds specifically to the MEF2 element present in the regulatory regions of many muscle-specific genes. Controls cardiac morphogenesis and myogenesis, and is also involved in vascular development.

MYC::MAX	v-*myc* avian myelocytomatosis viral oncogene homolog *MYC* associated factor X	The protein encoded by this gene is a multifunctional, nuclear phosphoprotein that plays a role in cell cycle progression, apoptosis and cellular transformation. It functions as a transcription factor that regulates transcription of specific target genes. The MYC:MAX complex is a transcriptional activator.
MIZF	Histone H4 transcription factor	MIZF interacts with methyl-CpG-binding protein-2 and plays a role in DNA methylation and transcription repression.
MYB	Myb proto-oncogene protein	This gene encodes a transcription factor that is a member of the MYB family of transcription factor genes. Transcriptional activator and plays an important role in the control of proliferation and differentiation of hematopoietic progenitor cells.
MZF1_1-4 MZF1_1-5-13	Myeloid zinc finger 1	Binds to target promoter DNA and functions as trancription regulator. May be one regulator of transcriptional events during hemopoietic development. Isoforms of this protein have been shown to exist at protein level.
NFATC2	Nuclear factor of activated T-cells, cytoplasmic 2	This protein is present in the cytosol and only translocates to the nucleus upon T cell receptor (TCR) stimulation, where it becomes a member of the nuclear factors of activated T cells transcription complex.
NFIC	Nuclear factor 1 C-type	Recognizes and binds the palindromic sequence 5'-TTGGCNNNNNGCCAA-3' present in viral and cellular promoters and in the origin of replication of adenovirus type 2. These proteins are individually capable of activating transcription and replication.
NFE2::MAF	Nuclear factor, erythroid 2 V-*maf* avian musculoaponeurotic fibrosarcoma oncogene homolog	Regulates erythroid and megakaryocytic maturation and differentiation. Plays a role in all aspects of hemoglobin production from globin and heme synthesis to procurement of iron. When overexpressed, represses anti-oxidant response element (ARE)-mediated transcription.
NFE2L1:MAFG	Nuclear factor erythroid 2-related factor 1 Transcription factor MafG	Nuclear factor erythroid 2-related factor (Nrf2) coordinates the up-regulation of cytoprotective genes via the antioxidant response element (ARE). MafG is a ubiquitously expressed small maf protein that is involved in cell differentiation of erythrocytes. It dimerizes with P45 NF-E2 protein and activates expression of a and b-globin.
NFKB1	Nuclear factor of kappa light polypeptide gene enhancer in B-cells 1	NF-kappa-B is a pleiotropic transcription factor present in almost all cell types and is the endpoint of a series of signal transduction events that are initiated by a vast array of stimuli related to many biological processes such as inflammation, immunity, differentiation, cell growth, tumorigenesis and apoptosis.
NFYB	Nuclear transcription factor Y, beta	The protein encoded by this gene is one subunit of a trimeric complex, forming a highly conserved transcription factor that binds with high specificity to CCAAT motifs in the promoter regions in a variety of genes. This gene product, subunit B, forms a tight dimer with the C subunit, a prerequisite for subunit A association. The resulting trimer binds to DNA with high specificity and affinity. Subunits B and C each contain a histone-like motif.
NHLH1	Nescient helix loop helix 1	The helix-loop-helix (HLH) proteins are a family of putative transcription factors, some of which have been shown to play an important role in growth and development of a wide variety of tissues and species.
NKX3-2	Natural killer 3 homeobox 2	This gene encodes a member of the NK family of homeobox-containing proteins. Transcriptional repressor that acts as a negative regulator of chondrocyte maturation.
NR2C2	Nuclear receptor subfamily 2, group C, member 2	Orphan nuclear receptor that can act as a repressor or activator of transcription. An important repressor of nuclear receptor signaling pathways such as retinoic acid receptor, retinoid X, vitamin D3 receptor, thyroid hormone receptor and estrogen receptor pathways.
NR2F1 (COUP)	Nuclear receptor subfamily 2, group F, member 1	Binds to the ovalbumin promoter and, in conjunction with another protein (S300-II) stimulates initiation of transcription. Binds to both direct repeats and palindromes of the 5'-AGGTCA-3' motif. An important paralog of this gene is RXRA.
NR4A2	Nuclear receptor subfamily 4, group A, member 2	Transcriptional regulator which is important for the differentiation and maintenance of meso-diencephalic dopaminergic (mdDA) neurons during development.
NRF1	Nuclear respiratory factor 1	This gene encodes a protein that homodimerizes and functions as a transcription factor which activates the expression of some key metabolic genes regulating cellular growth and nuclear genes required for respiration, heme biosynthesis, and mitochondrial DNA transcription and replication.
PAX2	Paired box gene 2	Probable transcription factor that may have a role in kidney cell differentiation.
PAX5	Paired box 5 gene	The central feature of this transcription factor gene family is the conserved DNA-binding paired box domain. Alternative splicing of this gene results in multiple transcript variants.
PLAG1	Pleomorphic adenoma gene 1	A zinc finger protein with 2 putative nuclear localization signals. PLAG1, which is developmentally regulated, has been shown to be consistently rearranged in pleomorphic adenomas of the salivary glands.
POU2F2	POU class 2 homeobox 2	The protein encoded by this gene is a homeobox-containing transcription factor of the POU domain family. The encoded protein binds the octamer sequence 5'-ATTTGCAT-3', a common transcription factor binding site in immunoglobulin gene promoters.
PPARA	Peroxisome proliferator-activated receptor alpha	Peroxisome proliferator-activated receptor alpha (PPARalpha) is a member of the nuclear receptor family of ligand-activated transcription factors that heterodimerize with the retinoic X receptor (RXR) to regulate gene expression. PPARalpha is located primarily in the liver, adipose tissue, kidney, heart, skeletal muscle and large intestine where it is thought to regulate fatty acid synthesis and oxidation, gluconeogenesis, ketogenesis and lipoprotein assembly.

PPARG:RXRA	Peroxisome proliferator-activated receptor gamma Retinoid X receptor, alpha	Peroxisome proliferator-activated receptor gamma (PPARgamma) is a member of the nuclear receptor family of ligand-activated transcription factors that heterodimerize with the retinoic X receptor (RXR) to regulate gene expression. PPARgamma is located primarily in the adipose tissue, lymphoid tissue, colon, liver and heart and is thought to regulate adipocyte differentiation and glucose homeostasis.
PRDM1	PR domain containing 1, with ZNF domain	This gene encodes a protein that acts as a repressor of beta-interferon gene expression.
RAR	Retinoic acid receptor, alpha	This gene represents a nuclear retinoic acid receptor. The encoded protein, retinoic acid receptor alpha, regulates transcription in a ligand-dependent manner. This gene has been implicated in regulation of development, differentiation, apoptosis, granulopoeisis, and transcription of clock genes.
REL	v-*rel* avian reticuloendotheliosis viral oncogene homolog	Proto-oncogene that may play a role in differentiation and lymphopoiesis. NF-kappa-B is a pleiotropic transcription factor which is present in almost all cell types and is involved in many biological processed such as inflammation, immunity, differentiation, cell growth, tumorigenesis and apoptosis.
RELA	v-rel avian reticuloendotheliosis viral oncogene homolog A	NF-kappa-B is a homo- or heterodimeric complex formed by the Rel-like domain-containing proteins **RELA**/p65, RELB, NFKB1/p105, NFKB1/p50, REL and NFKB2/p52 and the heterodimeric p65-p50 complex appears to be most abundant one.
RFX5	Regulatory factor X, 5	Activates transcription from class II MHC promoters. Recognizes X-boxes. Mediates cooperative binding between RFX and NF-Y. RFX binds the X1 box of MHC-II promoters
RORA_1 & 2	RAR-related orphan receptor A	Orphan nuclear receptor. Binds DNA as a monomer to hormone response elements (HRE) containing a single core motif half-site preceded by a short A-T-rich sequence. It has been shown to aid in the transcriptional regulation of some genes involved in circadian rhythm. Regulates a number of genes involved in lipid metabolism, in cerebellum and photoreceptor development and and skeletal muscle development.
RREB1	Ras responsive element binding protein	Transcription factor that binds specifically to the RAS-responsive elements (RRE) of gene promoters. May be involved in Ras/Raf-mediated cell differentiation by enhancing calcitonin expression. Represses the angiotensinogen gene. Negatively regulates the transcriptional activity of AR
RUNX1	Runt-related transcription factor 1	Heterodimeric transcription factor that binds to the core element of many enhancers and promoters. The protein encoded by this gene represents the alpha subunit of core binding factor and is thought to be involved in the development of normal hematopoiesis.
RUNX2	Runt-related transcription factor 2	Transcription factor involved in osteoblastic differentiation and skeletal morphogenesis. Essential for the maturation of osteoblasts and both intramembranous and endochondral ossification.
SOX9	SRY (sex determining region Y)-box 9	The protein encoded by this gene recognizes the sequence CCTTGAG along with other members of the
SOX10	SRY (sex determining region Y)-box 10	This gene encodes a member of the SOX (SRY-related HMG-box) family of transcription factors involved in the regulation of embryonic development and in the determination of the cell fate.
SP1	Specificity Protein 1	Can activate or repress transcription in response to physiological and pathological stimuli. Regulates the expression of a large number of genes involved in a variety of processes such as cell growth, apoptosis, differentiation and immune responses.
SP2	Specificity Protein 2	This gene encodes a member of the Sp subfamily of Sp/XKLF transcription factors. Sp family proteins are sequence-specific DNA-binding proteins characterized by an amino-terminal trans-activation domain and three carboxy-terminal zinc finger motifs. This protein contains the least conserved DNA-binding domain within the Sp subfamily of proteins, and its DNA sequence specificity differs from the other Sp proteins. It localizes primarily within subnuclear foci associated with the nuclear matrix, and can activate or in some cases repress expression from different promoters.
SPIB	Transcription factor Spi-B	SPI1 and SPIB are members of a subfamily of ETS transcription factors. ETS proteins share a conserved ETS domain that mediates specific DNA binding. SPIB and SPI1 bind to a purine-rich sequence, the PU box (5-prime-GAGGAA-3-).
SPL1	Squamosa promoter-binding-like protein 1	A member of SPL gene family, encodes DNA binding proteins and putative transcription factors.
SREBF1	Sterol regulatory element binding transcription factor 1	Transcriptional activator required for lipid homeostasis. Regulates transcription of the LDL receptor gene as well as the fatty acid and to a lesser degree the cholesterol synthesis pathway.
SREBF2	Sterol regulatory element binding transcription factor 2	Transcriptional activator required for lipid homeostasis. Regulates transcription of the LDL receptor gene as well as the fatty acid and to a lesser degree the cholesterol synthesis pathway.
SRY	Sex determining region Y	Transcriptional regulator that controls a genetic switch in male development.
STAT3	Signal transducer and activator of transcription 3 (acute-phase response factor)	Signal transducer and transcription activator that mediates cellular responses to interleukins, KITLG/SCF and other growth factors

STAT4	Signal transducer and activator of transcription 4	Carries out a dual function: signal transduction and activation of transcription. Involved in IL12 signaling.
STAT2::STAT1	Signal transducer and activator of transcription 2 Signal transducer and activator of transcription 1	The protein encoded by this gene is a member of the STAT protein family. In response to cytokines and growth factors, STAT family members are phosphorylated by the receptor associated kinases, and then form homo- or heterodimers that translocate to the cell nucleus where they act as transcription activators.
STAT5A:STAT5B	Signal transducer and activator of transcription 5A and transcription 5B	Carries out a dual function: signal transduction and activation of transcription. Regulates the expression of milk proteins during lactation.
TAL1::GATA1	T-cell acute lymphocytic leukemia 1: GATA binding protein 1 (globin transcription factor 1)	Implicated in the genesis of hemopoietic malignancies. It may play an important role in hemopoietic differentiation. Serves as a positive regulator of erythroid differentiation. GATA1 is a ranscriptional activator or repressor which probably serves as a general switch factor for erythroid development.
TAL1:TCF3	T-cell acute lymphocytic leukemia 1: Transcription factor 3	Implicated in the genesis of hemopoietic malignancies. It may play an important role in hemopoietic differentiation. Serves as a positive regulator of erythroid differentiation. TCF3 has been shown to directly enhance Hes1 (a well-known target of Notch signaling) expression.
TCF7L2	Transcription factor 7-like 2 (T-cell specific, HMG-box)	This gene encodes a high mobility group (HMG) box-containing transcription factor that plays a key role in the Wnt signaling pathway. The protein has been implicated in blood glucose homeostasis.
TCFCP211	Transcription factor CP2-like 1	Transcriptional suppressor. May suppress UBP1-mediated transcriptional activation. Modulates the placental expression of CYP11A1.
TFAP2a	Activator protein 2	The AP2a protein acts as a sequence specific DNA-binding transcription factor recognizing and binding to the specific DNA sequence and recruiting transcription machinery.
TFAP2c	Transcription factor AP-2 gamma	Sequence-specific DNA-binding protein that interacts with inducible viral and cellular enhancer elements to regulate transcription of selected genes. Activates genes involved in a large spectrum of important biological functions including proper eye, face, body wall, limb and neural tube development.
THAP1	THAP domain containing, apoptosis associated protein 1	DNA-binding transcription regulator that regulates endothelial cell proliferation and G1/S cell-cycle progression.
THR	Thyroid hormone receptor	The protein encoded by this gene is a nuclear hormone receptor for triiodothyronine. It is one of the several receptors for thyroid hormone, and has been shown to mediate the biological activities of thyroid hormone. Nuclear hormone receptor that can act as a repressor or activator of transcription. High affinity receptor for thyroid hormones, including triiodothyronine and thyroxine.
TP53	Tumor protein p53	A transcription factor whose protein levels and post-translational modification state alter in response to cellular stress (such as DNA damage, hypoxia, spindle damage).
USF1 & 2	Upstream transcription factor 1 & 2	This gene encodes a member of the basic helix-loop-helix leucine zipper family, and can function as a cellular transcription factor. The encoded protein can activate transcription through pyrimidine-rich initiator (Inr) elements and E-box motifs.
YY1	YY1 transcription factor	YY1 is a ubiquitously distributed transcription factor belonging to the GLI-Kruppel class of zinc finger proteins. The protein is involved in repressing and activating a diverse number of promoters. YY1 may direct histone deacetylases and histone acetyltransferases to a promoter in order to activate or repress the promoter, thus implicating histone modification in the function of YY1.
ZEB1	Zinc finger E-box-binding homeobox 1	A member of the delta-EF1 (TCF8)/Zfh1 family of 2-handed zinc finger/homeodomain proteins and interacts SMADs with receptor-mediated, activated full-length activated full-length SMADs.
ZFX	Zinc finger X-chromosomal protein	A member of the krueppel C2H2-type zinc-finger protein family and probable transcriptional activator.
ZNF263	Zinc finger protein 263	Might play an important role in basic cellular processes as a transcriptional repressor. An important paralog to ZNF496.
ZNF354C	Zinc finger protein 354C	May function as a transcription repressor.