OPEN ACCESS
Freely Available Online

JOURNAL OF PROTEOMICS AND GENOMICS RESEARCH
ISSN NO: 2326-0793

RESEARCH ARTICLE
DOI : 10.14302/issn.2326-0793.jpgr-15-889

Computational *EPAS1* rSNP Analysis, Transcriptional Factor Binding Sites and High Altitude Sickness or Adaptation

Norman E. Buroker[1,*]

1. Department of Pediatrics, University of Washington, Seattle, WA 98195, USA.

Abstract

Purpose – The endothetal Per-Arnt-Sim (PAS) domain protein 1 (*EPAS1*) gene which encodes hypoxia-inducible-factor-2 alpha (*HIF2a*) is a transcription factor that is involved in the response to hypoxia. *EPAS1* has been found to have four (rs56721780, rs6756667, rs7589621, rs1868092) simple nucleotide polymorphisms (SNPs) associated with human disease. These SNPs were computationally examined with respect to changes in potential transcriptional factor binding sites (TFBS) and these changes were discussed in relation to disease and alterations in high altitude adaptation in humans.

Methods -- The JASPAR CORE and ConSite databases were instrumental in identifying the TFBS. The Vector NTI Advance 11.5 computer program was employed in locating all the TFBS in the *EPAS1* gene from 1.6 kb upstream of the transcriptional start site to 539 bps past the 3'UTR. The JASPAR CORE database was also involved in computing each nucleotide occurrence (%) within the TFBS.

Results – The *EPAS1* SNPs in the promoter, intron two and the 3'UTR regions have previously been found to be significantly associated with disease and different levels of high-altitude hypoxia among native Tibetans. The SNP alleles were found to alter the DNA landscape for potential transcriptional factors (TFs) to attach resulting in changes in TFBS and thereby, alter which transcriptional factors potentially regulate the *EPAS1* gene such as for the glucocorticoid and mineralocorticoid nuclear receptor binding sites created by the rs7589621 rSNP *EPAS1*-G allele. These receptors regulate carbohydrate, protein and fat metabolism. Also the minor rs7589621 rSNP *EPAS1*-A creates a punitive TFBS for the FOXC TF which is an important regulator of cell viability and resistance to oxidative stress. These *EPAS1* SNPs should be considered as regulatory (r) SNPs.

Conclusion -- The alleles of each rSNP were found to generate unique TFBS resulting in potential changes in TF *EPAS1* regulation. The punitive changes in TFBS created by the four rSNPs could very well influence the significant cline in allele frequencies seen in Tibetans with increasing altitude or the haplotype association with high altitude polycythemia in male Han Chinese. These regulatory changes were discussed with respect to changes in human health that result in disease and sickness.

OPEN ACCESS
Freely Available Online

OpenAccessPub

Corresponding author:

Norman E. Buroker, Department of Pediatrics, University of Washington, Seattle, WA 98195, USA.
E-mail: nburoker@u.washington.edu

Keywords: *EPAS1*, rSNPs, TFBS, disease, sickness

Received: Dec 15, 2015; **Accepted**: Feb 07, 2016; **Published** : Feb 20, 2016;

Introduction:

The endothetal Per-Arnt-Sim (PAS) domain protein 1 (*EPAS1*) gene which encodes hypoxia-inducible -factor-2 alpha (*HIF2a*) is a transcription factor that is involved in the response to hypoxia. Hypoxia is a major geographical condition associated with high-altitude environments [1]. Hypoxia-inducible-factors (HIFs) are heterodimers consisting of an oxygen-labile HIFa subunit and a stable HIFb subunit [2]. During hypoxia conditions, three isoforms either *HIF1a, HIF2a or HIF3a* and *HIF1b* are activated and function as transcriptional regulators of genes involved with the hypoxia response [3-5]. Genome wide association studies (GWAS) on high -altitude adaptation have implicated several single nucleotide polymorphisms (SNPs) in the regulatory region of the *EPAS1* gene which are responsible for the genetic adaptation of high-altitude hypoxia in Tibetans [6-8]. Genetic variation in the regulatory region of the *EPAS1* gene may influence gene expression and contribute to changes in biological functions [9]. *EPAS1* is expressed in organs that are involved in oxygen transport and metabolism, such the lung, placenta and vascular endothelium [10], and is associated with many biological processes and diseases related to metabolism [11], angiogenesis [12, 13], inflammation [14, 15] and cancer [16-18].

The *EPAS1* gene maps to human chromosome 2p21 and is about 120 kb in size with a coding region consisting of 15 exons [19]. Four *HIF2a* SNPs (rs56721780, rs6756667, rs7589621 and rs1868092) have been significantly associated with different levels of high-altitude hypoxia among native Tibetans [20]. The rs56721780 SNP in the *HIF2a* promoter region has also been significantly associated with high-altitude adaptation of Tibetans [9] while the rs6756667 SNP from intron two has been significantly associated with susceptibility to acute mountain sickness in individuals unaccustomed to high altitude environments [21]. The rs1868092 SNP near the *HIF2a* 3'UTR has been associated with high altitude polycythemia in male Han Chinese at the Qinghai-Tibetan plateau [22].

Single nucleotide changes that affect gene expression by impacting gene regulatory sequences such as promoters, enhances, and silencers are known as regulatory SNPs (rSNPs) [23-26]. A rSNPs within a transcriptional factor binding site (TFBS) can change a transcriptional factor's (TF) ability to bind its TFBS [27-30] in which case the TF would be unable to effectively regulate its target gene [31-35]. This concept is examined for the above four *HIF2a* rSNPs and their allelic association with TFBS, where computation analyses [36-39] is used to identify TFBS alterations created by the *HIF2a* rSNPs. In this report, the rSNP associations with changes in potential TFBS are discussed with their possible relationship to disease or sickness in humans.

Methods

The JASPAR CORE database [40, 41] and ConSite [42] were used to identify the potential *STAT4* TFBS in this study. JASPAR is a database 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 rSNP location within the binding sites have previously been discussed [43]. The Vector NTI Advance 11.5 computer program (Invitrogen, Life Technologies) was used to locate the TFBS in the *EPAS1* gene (NCBI Ref Seq NM_001430) from 1.6 kb upstream of the transcriptional start site to 539 bps past the 3'UTR which represents a total of 91 kb. The JASPAR CORE database was also used to calculate each nucleotide occurrence (%) within the TFBS, where upper case lettering indicate that the nucleotide occurs 90% or greater and lower case less than 90%. The occurrence of each SNP allele in the TFBS is also computed from the database (Table 3).

Results

EPAS1 rSNPs and TFBS

The allele frequencies of four *EPAS1* SNPs (rs56721780, rs6756667, rs7589621 and rs1868092) significantly associated with different levels of high-altitude hypoxia among native Tibetans [20] are presented in Table 1 along with low altitude Han Chinese and Japanese populations. The common rs56721780 SNP *EPAS1* -G allele creates nine unique punitive TFBS for the REL, RELA, RUNX1,TFAP2A, TFAP1 (var.2), TFAP2B, TFAP2B(var.2), TFAP2C and TFAP2C (var.2) TFs, which are involved with inflammation, immunity, differentiation, cell growth, tumorigenesis, apoptosis, hematopoiesis, transcriptional activation and repression, respectively (Tables 2 & 3). The minor *EPAS1* -C allele creates two unique punitive TFBS for the FOXP3 and HOXA5 TFs which are involved with the homeostasis of the immune system and specific identities on the anterior-posterior axis during development, respectively (Tables 2 & 3). There are also four conserved TBFS for the HLTF, HNF4G, NFAT5 and SOX10 TFs which are involved altering chromatin structure, transcription, regulation of osmoprotective and inflammatory genes and embryonic development, respectively (Table 2, Table 3).

The common rs6756667 SNP *EPAS1*-A allele creates thirteen unique punitive TBFS for the CEBPa, FIGLA, FOS::JUN, FOXH1, HIC1 & 2, HOXA2, NFIA, NFIC, NRL, RHOXFA, TEAD3 and THAP1 TFs, which are involved with enhancers, folliculogenesis, signal transduction, tumor suppression, cell specific positional identities, transcription and replication, photoceptor development and function, transcriptional enhancer and transcription regulation respectively (Tables 2 & 3). The minor *EPAS1* -G allele creates nine unique punitive TFBS for the ATF7, GMEB2, INSM1, JDP2 (var.2), MGA, NR4A2, USF1 & 2, and ZNF354C TFs which are involved with early cell signaling, DNA replication, neuroendocrine differentiation of human lung tumors, tumorigenesis and anti-tumorigenesis, transcription activator or repressor, transcription regulator, cellular transcriptional factor and transcription repression, respectively (Table 3, Table 2). There are also twenty-one conserved TFBS for the ATF4, CEBPb, CEBPd, CEBPe, CREB1, DBP, HLF, JUN, JUND (var.2), MEIS1 & 3, NFE2L1:: MAFG, NFIX, RUNX1 & 3, SREBF1 & 2, SREBF2 (var.2), TBX4 & 5 and TFEC TFs which are c-AMP-response element binding proteins, DNA-binding proteins, immune and inflammatory responses, circadian rhythm, transcription activation, signaling pathway stimulator and enhancer, normal development, up-regulation of cytoprotective genes via the antioxidant response element, enhancer sequence-specific DNA binding TF, development of normal hematopoiesis, tumor suppressor, lipid homeostasis, cholesterol homeostasis, mesoderm differentiation and cellular processes, respectively (Tables 2 & 3).

The common rs7589621 SNP *EPAS1*-G allele creates ten unique punitive TFBS for the EN1, FOXA1, GBX2, MEOX1 & 2, MIXL1, MSX1, NR3C1, NR3C2 and PHOX2A TFs which are involved with controlling development, embryonic development, cell pluripotency, sclerotome development, transcriptional repressor,

Table 1. *EPAS1 (HIF2a)* SNPs and high altitude hypoxia among native Tibetans. These SNPs have been found to be significantly associated with hypoxia in Tibetan populations. The SNPs are located in the EPAS1 gene. MAF is the minor allele frequency. Allele frequency data from reference 20.

					MAF Tibetan populations/Altitude (M)				Han Chinese	Japanese
Gene *EPAS1 (HIF2a)*	Gene Position	SNP	Chr Pos	Alleles	Bomi/2700	Qamdo/3200	Lhasa/3700	Amdo/4700	CHB/Beijing	JPT/Tokyo
	promoter	rs56721780	2:46523655	G/C	C=0.184	C=0.159	C=0.328	C=0.311	C=0.01	C=0.022
	intron 2	rs6756667	2:46579409	A/G	G=0.313	G=0.266	G=0.201	G=0.104	G=0.944	G=0.889
	intron 2	rs7589621	2:46582382	G/A	A=0.254	A=0.221	A=0.163	A=0.079	A=0.789	A=0.767
	past 3'UTR	rs1868092	2:46614202	A/G	G=0.359	G=0.327	G=0.249	G=0.156	G=0.919	G=0.924

Table 2. Transcriptional factor (TF) abbreviations, protein name and descriptions.

TFs	Protein name	TF description
ATF4	Activating Transcription Factor 4	The protein encoded by this gene belongs to a family of DNA-binding proteins that includes the AP-1 family of transcription factors, cAMP-response element binding proteins (CREBs) and CREB-like proteins.
ATF7	Activating Transcription Factor 7	Plays important functions in early cell signaling.Has no intrinsic transcriptional activity, but activates transcription on formation of JUN or FOS heterodimers.
BARX1	BARX Homeobox 1	Transcription factor, which is involved in craniofacial development, in odontogenesis and in stomach organogenesis.
BSX	Brain-Specific Homeobox	DNA binding protein that function as transcriptional activator. Is essentiel for normal post-natal growth and nursing. Is an essential factor for neuronal neuropeptide Y and agouti-related peptide function and locomotory behavior in the control of energy balance.
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. Binds to regulatory regions of several acute-phase and cytokines genes and probably plays a role in the regulation of acute-phase reaction, inflammation and hemopoiesis.
CEBPg	CCAAT/enhancer binding protein (C/EBP), delta	The encoded protein is important in the regulation of genes involved in immune and inflammatory responses, and may be involved in the regulation of genes associated with activation and/or differentiation of macrophages.
CEBPe	CCAAT/enhancer binding protein (C/EBP), epsilon	The encoded protein may be essential for terminal differentiation and functional maturation of committed granulocyte progenitor cells. Mutations in this gene have been associated with Specific Granule Deficiency, a rare congenital disorder.
CREB1	CAMP Responsive Element Binding Protein 1	This gene encodes a transcription factor that is a member of the leucine zipper family of DNA binding proteins
DBP	D Site Of Albumin Promoter (Albumin D-Box) Binding Protein	The encoded protein can bind DNA as a homo- or heterodimer and is involved in the regulation of some circadian rhythm genes.
DLX6	Distal-Less Homeobox 6	This gene encodes a member of a homeobox transcription factor gene family similiar to the Drosophila distal-less gene. This family is comprised of at least 6 different members that encode proteins with roles in forebrain and craniofacial development.
E2F6	E2F transcription factor 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
EN1	Engrailed homeobox 1	Homeobox-containing genes are thought to have a role in controlling development.
EN2	Engrailed homeobox 2	The human engrailed homologs 1 and 2 encode homeodomain-containing proteins and have been implicated in the control of pattern formation during development of the central nervous system.
ESX1	ESX Homeobox 1	This gene likely plays a role in placental development and spermatogenesis.
EVX1	Even-Skipped Homeobox 1	May play a role in the specification of neuronal cell types
EVX2	Even-Skipped Homeobox 2	The encoded protein is a homeobox transcription factor that is related to the protein encoded by the Drosophila even-skipped (eve) gene, a member of the pair-rule class of segmentation genes.
FIGLA	Folliculogenesis Specific Basic Helix-Loop-Helix	The protein is a basic helix-loop-helix transcription factor that regulates multiple oocyte-specific genes, including genes involved in folliculogenesis and those that encode the zona pellucida.
FOS	Jun Proto-Oncogene	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. The FOS proteins have been implicated as regulators of cell proliferation, differentiation, and transformation. Controls osteoclast survival and size. As a dimer with JUN, activates LIF transcription.

Table 2 (Continuation)

TFs	Protein name	TF description
FOS::JUN	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.
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.
FOXC1	Forkhead box C1	This gene belongs to the forkhead family of transcription factors which is characterized by a distinct DNA-binding forkhead domain. An important regulator of cell viability and resistance to oxidative stress.
FOXH1	Forkhead Box H1	Transcriptional activator. Recognizes and binds to the DNA sequence 5-TGT[GT][GT]ATT-3. Required for induction of the goosecoid (GSC) promoter by TGF-beta or activin signaling.
FOXP3	Forkhead Box P3	Transcriptional regulator which is crucial for the development and inhibitory function of regulatory T-cells (Treg). Plays an essential role in maintaining homeostasis of the immune system by allowing the acquisition of full conventional T-cells. Suppressive function and stability of the Treg lineage, and by directly modulating the expansion and function of conventional T-cells.
GBX1	Gastrulation Brain Homeobox 1	Sequence-specific DNA binding transcription factor activity and sequence-specific DNA binding. An important paralog of this gene is DLX5.
GBX2	Gastrulation Brain Homeobox 2	May act as a transcription factor for cell pluripotency and differentiation in the embryo
GMEB2	Glucocorticoid Modulatory Element Binding Protein 2	This gene is a member of KDWK gene family. The product of this gene associates with GMEB1 protein, and the complex is essential for parvovirus DNA replication.
GSX1	GS Homeobox 1	Activates the transcription of the GHRH gene. Plays an important role in pituitary development.
HIC1	Hypermethylated In Cancer 1	This gene functions as a growth regulatory and tumor repressor gene.
HIC2	Hypermethylated In Cancer 2	Transcriptional repressor
HLF	Hepatic Leukemia Factor	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	This gene encodes a member of the SWI/SNF family. Members of this family have helicase and ATPase activities and are thought to regulate transcription of certain genes by altering the chromatin structure around those genes.
HMBOX1	Homeobox Containing 1	Transcription factor. Isoform 1 acts as a transcriptional repressor.
HNF4g	Hepatocyte Nuclear Factor 4, Gamma	Transcription factor. Has a lower transcription activation potential than HNF4-alpha
HOXA2	Homeobox A2	Sequence-specific transcription factor which is part of a developmental regulatory system that provides cells with specific positional identities on the anterior-posterior axis.
HOXA5	Hoxa5	Sequence-specific transcription factor which is part of a developmental regulatory system that provides cells with specific positional identities on the anterior-posterior axis.
HOXB2	Homeobox B2	Sequence-specific transcription factor which is part of a developmental regulatory system that provides cells with specific positional identities on the anterior-posterior axis.
HOXB3	Homeobox B3	The encoded protein functions as a sequence-specific transcription factor that is involved in development.
INSM1	Insulinoma-Associated 1	This gene is a sensitive marker for neuroendocrine differentiation of human lung tumors.
ISL2	ISL LIM Homeobox 2	Transcriptional factor that defines subclasses of motoneurons that segregate into columns in the spinal cord and select distinct axon pathways.
ISX	Intestine-Specific Homeobox	Transcription factor that regulates gene expression in intestine. May participate in vitamin A metabolism most likely by regulating BCO1 expression in the intestine.
JDP(var.2)	Jun Dimerization Protein 2	Component of the AP-1 transcription factor that represses transactivation mediated by the Jun family of proteins. Involved in a variety of transcriptional responses associated with AP-1 such as UV-induced apoptosis, cell differentiation, tumorigenesis and antitumogeneris.
JUN	Jun Proto-Oncogene	Transcription factor that recognizes and binds to the enhancer heptamer motif 5-TGA[CG]TCA-3. signaling pathway stimulation. Promotes activity of NR5A1 when phosphorylated by HIPK3 leading to increased steroidogenic gene expression upon cAMP signaling pathway

Table 2 (Continuation)

TFs	Protein name	TF description
JUND(var.2)	Jun D Proto-Oncogene	Transcription factor that recognizes and binds to the enhancer heptamer motif 5'-TGA[CG]TCA-3'.
KLF5	Kruppel-like factor 5 (intestinal)	Transcription factor that binds to GC box promoter elements. Activates transcription of genes.
LBX2	Ladybird Homeobox 2	Putative transcription factor.
LIN54	Lin-54 DREAM MuvB Core Complex Component	Is a component of the LIN, or DREAM, complex, an essential regulator of cell cycle genes
MAX	MGA, MAX Dimerization Protein	The protein encoded by this gene is a member of the basic helix-loop-helix leucine zipper (bHLHZ) family of transcription factors
MEIS1	Meis Homeobox 1	Homeobox genes, of which the most well-characterized category is represented by the HOX genes, play a crucial role in normal development.
MEIS3	Meis Homeobox 3	Sequence-specific DNA binding and RNA polymerase II core promoter proximal region sequence-specific DNA binding transcription factor activity involved in positive regulation of transcription.
MEOX1	Mesenchyme Homeobox 1	Mesodermal transcription factor that plays a key role in somitogenesis and is specifically required for sclerotome development.
MEOX2	Mesenchyme Homeobox 2	The encoded protein may play a role in the regulation of vertebrate limb myogenesis. Mutations in the related mouse protein may be associated with craniofacial and/or skeletal abnormalities, in addition to neurovascular dysfunction observed in Alzheimer's disease.
MGA	MGA, MAX Dimerization Protein	Functions as a dual-specificity transcription factor, regulating the expression of both MAX-network and T-box family target genes. Functions as a repressor or an activator.
MIXL1	Mix Paired-Like Homeobox	Regulates cell fate during development.
MSX1	Msh Homeobox 1	Acts as a transcriptional repressor. May play a role in limb-pattern formation. Acts in cranofacial development and specifically in odontogenesis.
MZF1	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.
NEUROD2	Neuronal Differentiation 2	Transcriptional regulator implicated in neuronal determination. Mediates calcium-dependent transcription activation by binding to E box-containing promoter. Critical factor essential for the repression of the genetic program for neuronal differentiation; prevents the formation of synaptic vesicle clustering at active zone to the presynaptic membrane in postmitotic neurons.
NFAT5	Nuclear Factor Of Activated T-Cells 5, Tonicity-Responsive	Transcription factor involved in the transcriptional regulation of osmoprotective and inflammatory genes. Regulates hypertonicity-induced cellular accumulation of osmolytes.
NFATC3	Nuclear Factor Of Activated T-Cells, Cytoplasmic, Calcineurin-Dependent 3.	Acts as a regulator of transcriptional activation. Plays a role in the inducible expression of cytokine genes in T-cells, especially in the induction of the IL-2.
NFE2L1:MAFG	Nuclear Factor, Erythroid 2-Like 1 V-Maf Avian Musculoaponeurotic Fibrosarcoma Oncogene Homolog G	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.
NFIA	Nuclear Factor I/A	Recognizes and binds the palindromic sequence 5-TTGGCNNNNNGCCAA-3 present in viral and cellular promoters transcription and replication and in the origin of replication of adenovirus type 2. These proteins are individually capable of activating transcription and replication
NFIC	Nuclear Factor I/C (CCAAT-Binding Transcription Factor)	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.
NFIL3	Nuclear factor, interleukin 3 regulated	Expression of interleukin-3 (IL3; MIM 147740) is restricted to activated T cells, natural killer (NK) cells, and mast cell lines.

Table 2 (Continuation)

TFs	Protein name	TF description
NFIX	Nuclear Factor I/X (CCAAT-Binding Transcription Factor)	Sequence-specific DNA binding transcription factor activity and RNA polymerase II distal enhancer sequence-specific DNA binding transcription factor activity.
NKX2-3	NK2 Homeobox 3	This gene encodes a homeodomain-containing transcription factor. The encoded protein is a member of the NKX family of homeodomain transcription factors.
NKX2-8	NK2 Homeobox 8	Transcriptional factor. Diseases associated with NKX2-8 include esophageal cancer.
NKX3-1	NK3 Homeobox 1	This gene encodes a homeobox-containing transcription factor. This transcription factor functions as a negative regulator of epithelial cell growth in prostate tissue.
NKX3-2	NK3 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.
NKX6-1	NK6 Homeobox 1	Transcription factor which binds to specific A/T-rich DNA sequences in the promoter regions of a number of genes. Involved in transcriptional regulation in islet beta cells. Binds to the insulin promoter and is involved in regulation of the insulin gene.
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.
NR3C1	Nuclear Receptor Subfamily 3, Group C, Member 1 (Glucocorticoid Receptor)	Glucocorticoids regulate carbohydrate, protein and fat metabolism, modulate immune re-sponses through supression of chemokine and cytokine production and have critical roles in constitutive activity of the CNS, digestive, hematopoietic, renal and reproductive systems. The protein encoded by this gene plays a role in protecting cells from oxidative stress and damage induced by ionizing radiation.
NR3C2	Nuclear Receptor Subfamily 3, Group C, Member 2	This gene encodes the mineralocorticoid receptor, which mediates aldosterone actions on salt and water balance within restricted target cells.
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.
NRL	Neural Retina Leucine Zipper	This gene encodes a basic motif-leucine zipper transcription factor of the Maf subfamily. The encoded protein is conserved among vertebrates and is a critical intrinsic regulator of photoceptor development and function.
PDX1	Pancreatic and duodenal homeobox 1	Activates insulin, somatostatin, glucokinase, islet amyloid polypeptide and glucose transporter type 2 gene transcription. Particularly involved in glucose-dependent regulation of insulin gene transcription.
PHOX2A	Paired-Like Homeobox 2a	May be involved in regulating the specificity of expression of the catecholamine biosynthetic genes. Acts as a transcription activator/factor.
POU2F1	POU Class 2 Homeobox 1	Transcription factor that binds to the octamer motif (5-ATTTGCAT-3) and activates the promoters of the genes for some small nuclear RNAs (snRNA) and of genes such as those for histone H2B and immunoglobulins. Modulates transcriptiontransactivation by NR3C1, AR and PGR
POU3F1	POU Class 3 Homeobox 1	Transcription factor that binds to the octamer motif (5-ATTTGCAT-3). Thought to be involved in early embryogenesis and neurogenesis
POU3F2	POU Class 3 Homeobox 2	This gene encodes a member of the POU-III class of neural transcription factors. The en-coded protein is involved in neuronal differentiation and enhances the activation of cortico-tropin-releasing hormone regulated genes.
POU3F3	POU Class 3 Homeobox 3	This gene encodes a POU-domain containing protein that functions as a transcription factor. The encoded protein recognizes an octamer sequence in the DNA of target genes. This protein may play a role in development of the nervous system.
POU3F4	POU Class 3 Homeobox 4	This gene encodes a member of the POU-III class of neural transcription factors. This family member plays a role in inner ear development. The protein is thought to be involved in the mediation of epigenetic signals which induce striatal neuron-precursor differentiation.
POU5F1	POU Class 5 Homeobox 1B	This gene encodes a transcription factor containing a POU homeodomain that plays a key role in embryonic development and stem cell pluripotency. Aberrant expression of this gene in adult tissues is associated with tumorigenesis. Forms a trimeric complex with SOX2 on DNA and controls the expression of a number of genes involved in embryonic development such as YES1, FGF4, UTF1 and ZFP206.

Table 2 (Continuation)

TFs	Protein name	TF description
POU5F1B	POU Class 5 Homeobox 1B	This intronless gene was thought to be a transcribed pseudogene of POU class 5 homeobox 1; however, it has been reported that this gene can encode a functional protein. The protein has been shown to be a weak transcriptional activator and may play a role in carcinogenesis and eye development.
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 processes such as inflammation, immunity, differentiation, cell growth, tumorigenesis and apoptosis.
RELA	V-Rel Avian Reticuloendotheliosis Viral Oncogene Homolog A	RELA is a Protein Coding gene. NF-kappa-B is composed of NFKB1 or NFKB2 bound to either REL, RELA, or RELB. The most abundant form of NF-kappa-B is NFKB1 complexed with the product of this gene, RELA. Among its related pathways are PI3K-Akt signaling pathway and PI-3K cascade.
RHOXF1	Rhox Homeobox Family, Member 1	This gene is a member of the PEPP subfamily of paired-like homoebox genes. The gene may be regulated by androgens and epigenetic mechanisms. The encoded nuclear protein is likely a transcription factor that may play a role in human reproduction.
RUNX1	Runt-Related Transcription Factor 1	Core binding factor (CBF) is a 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 CBF and is thought to be involved in the development of normal hematopoiesis.
RUNX3	Runt-Related Transcription Factor 3	This gene encodes a member of the runt domain-containing family of transcription factors. Found in a number of enhancers and promoters, and can either activate or suppress transcription. It also interacts with other transcription factors. It functions as a tumor suppressor, and the gene is frequently deleted or transcriptionally silenced in cancer.
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.
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	This gene encodes a member of the a ubiquitously expressed transcription factor that controls cholesterol homeostasis by regulating transcription of sterol-regulated genes. The encoded protein contains a basic helix-loop-helix-leucine zipper (bHLH-Zip) domain and binds the sterol regulatory element 1 motif.
SRY	SRY (sex determining region Y)-box 10	Transcriptional regulator that controls a genetic switch in male development. It is necessary and sufficient for initiating male sex determination by directing the development of supporting cell precursors
TBP	TATA Box Binding Protein	General transcription factor that functions at the core of the DNA-binding multiprotein factor TFIID. Binding of TFIID to the TATA box is the initial transcriptional step of the pre-initiation complex (PIC), playing a role in the activation of eukaryotic genes transcribed by RNA polymerase II.
TBX4	T-Box 4	Involved in the transcriptional regulation of genes required for mesoderm differentiation.
TBX5	T-Box 5	This gene is a member of a phylogenetically conserved family of genes that share a common DNA-binding domain, the T-box. T-box genes encode transcription factors involved in the regulation of developmental processes.
TEAD1	TEA Domain Family Member 1	This gene encodes a ubiquitous transcriptional enhancer factor that is a member of the TEA/ATTS domain family. This protein directs the transactivation of a wide variety of genes and, in placental cells, also acts as a transcriptional repressor.
TEAD3	TEA Domain Family Member 3	This gene product is a member of the transcriptional enhancer factor (TEF) family of transcription factors, which contain the TEA/ATTS DNA-binding domain. It is predominantly expressed in the placenta and is involved in the transactivation of the chorionic somatomammotropin-B gene enhancer.
TEAD4	TEA Domain Family Member 4	It is preferentially expressed in the skeletal muscle, and binds to the M-CAT regulatory element found in promoters of muscle-specific genes to direct their gene expression.

Table 2 (Continuation)

TFs	Protein name	TF description
TFAP2A	Transcription Factor AP-2 Alpha (Activating Enhancer Binding Protein 2 Alpha)	The protein encoded by this gene is a transcription factor that binds the consensus sequence 5'-GCCNNNGGC-3' and activates the transcription of some genes while inhibiting the transcription of others.
TFAP2B	Transcription Factor AP-2 Beta (Activating Enhancer Binding Protein 2 Beta)	This gene encodes a member of the AP-2 family of transcription factors. AP-2 proteins form homo- or hetero-dimers with other AP-2 family members and bind specific DNA sequences. This protein functions as both a transcriptional activator and repressor.
TFAP2C	Transcription Factor AP-2 Gamma (Activating Enhancer Binding Protein 2 Gamma)	Sequence-specific DNA-binding protein that interacts with inducible viral and cellular enhancer elements to regulate transcription of selected genes. AP-2 factors bind to the consensus sequence 5'-GCCNNNGGC-3' and activate genes involved in a large spectrum of important biological functions including proper eye, face, body wall, limb and neural tube development.
TFEC	Transcription Factor EC	This gene encodes a member of the micropthalmia (MiT) family of basic helix-loop-helix leucine zipper differentiation. MiT transcription factors regulate the expression of target genes by binding to E-box recognition sequences as homo- or heterodimers, and play roles in multiple cellular processes including survival, growth and and differentiation.
THAP1	THAP domain containing, apoptosis associated protein 1	DNA-binding transcription regulator that regulates endothelial cell proliferation and G1/S cell -cycle progression.
USF1	Upstream Transcription Factor 1	This gene encodes a member of the basic helix-loop-helix leucine zipper family, and can function as a cellular.transcription factor.
USF2	Upstream Transcription Factor 2, C-Fos Interacting	Transcription factor that binds to a symmetrical DNA sequence (E-boxes) (5-CACGTG-3) that is found in a variety of viral and cellular promoters.
YY1	YY1 Transcription Factor	Multifunctional transcription factor that exhibits positive and negative control on a large number of cellular and viral genes by binding to sites overlapping the transcription start site
YY2	YY2 Transcription Factor	The protein encoded by this gene is a transcription factor that includes several Kruppel-like zinc fingers in its C-terminal region. It possesses both activation and repression domains, and it can therefore have both positive and negative effects on the transcription of target genes.
ZNF354C	Zinc finger protein 354C	May function as a transcription repressor.

Table 3. The *EPAS1 (HIF2a)* SNPs that were examined in this study where the minor allele is in red. Also listed are the transcriptional factors (TF), their potential binding sites (TFBS) containing these SNPs and DNA strand orientation. TFs in red differ between the SNP alleles. Where upper case nucleotide designates the 90% conserved BS region and red is the SNP location of the alleles in the TFBS. Below the TFBS is the nucleotide occurrence (%) obtained from the Jaspar Core database. Also listed are the number (#) of binding sites in the gene for the given TF. Note: TFs can bind to more than one nucleotide sequence.

EPAS1 (HIF2a)					
SNP	**Allele**	**TFs**	**# of Sites**	**TFBS**	**Strand**
rs56721780	**G**	HLTF	1	agcCtTtggg g=14%	plus
		HNF4G	1	gaaaccCAaAGgcta c=30%	minus
		NFAT5	1	ggTTtCccag g=21%	plus
		REL	1	tggttTccC g=6%	plus
		REL	1	ttggtTtcC g=53%	plus
		RELA	1	ttGgtTtCC g=39%	plus
		RELA	1	tgGgttTcCC G=100%	plus
		RUNX1	2	gccTttGGgtt G=92%	plus
		SOX10	76	cttTgg g=5%	plus
		TFAP2A	1	acCCaaagGct C=99%	minus
		TFAP2A	1	agCCtttgGgt G=99%	plus
		TFAP2A(var.2)	1	aaCCcaaaGgct C=99%	minus
		TFAP2A(var.2)	1	agCCtttgGgtt G=94%	plus
		TFAP2B	1	aaCCcaaaGgct C=99%	minus
		TFAP2B	1	agCCtttgGgtt G=97%	plus
		TFAP2B(var.2)	1	acCCaaaGGCt C=100%	minus
		TFAP2C	1	aacCcaaaGgct C=97%	minus
		TFAP2C	1	agcCtttgGgtt G=97%	plus
		TFAP2C(var.2)	1	acCCaaagGCt C=98%	minus
		TFAP2C(var.2)	1	agCCtttgGgt G=99%	plus

Table 3 (Continuation)

	C	FOXP3	20	gcaaAgg g=65%	minus
		HLTF	1	agcCtTtgcg c=24%	plus
		HNF4G	1	gaaacgCAaAGgcta g=24%	minus
		HOXA5	2	cgcaaagg g=31%	minus
		NFAT5	1	cgTTtCccag c=21%	plus
		SOX10	75	cttTgc c=0%	plus
rs6756667	**A**	ATF4	1	aggTGAtGccAca t=48%	minus
		CEBPα	1	gTggCatcAcc a=78%	plus
		CEBPβ	3	gTgatgccAc t=10%	minus
		CEBPβ	2	aTgccacAAt T=100%	minus
		CEBPδ	3	gTgatgccAc t=9%	minus
		CEBPδ	2	aTgccacAAt T=100%	minus
		CEBPε	3	gTgatgccAc t=18%	minus
		CEBPε	2	aTgccacAAt T=98%	minus
		CREB1	6	tGAtGcca t=0%	minus
		DBP	1	ggTgAtgccAca t=38%	minus
		DBP	1	tgTggcatcAcc a=45%	plus
		FIGLA	1	atCAcctTac a=50%	plus
		FOS::JUN	34	TgccacA T=94%	minus
		FOXH1	3	gtgAtgccACa t=5%	minus
		HIC1	2	aTgCCacaa T=95%	minus
		HIC2	2	aTgCCacaa T=98%	minus
		HLF	1	ggTgatgccaca t=11%	minus

Table 3 (Continuation)

		HLF	1	tgTggcatcacc a=33%	plus
		HOXA2	2	tggcATcAcc A=90%	plus
		JUN	1	aaggTGAtGccAc t=62%	minus
		JUND (var.2)	1	taaggTGAtgccAca t=44%	minus
		JUND (var.2)	1	ggtgaTGccacaAtc T=100%	minus
		MEIS1	15	atGcCac t=77%	minus
		MEIS1	15	gtGgCat a=83%	plus
		MEIS3	6	gtGgCAtc A=91%	plus
		NFE2L1::MafG	40	caTcAc a=85%	plus
		NFIA	1	gaTGCCAcaa T=100%	minus
		NFIC	84	gTGGca a=48%	plus
		NFIX	4	gatGCCAca t=18%	minus
		NFIX	4	tgtGgCAtc A=92%	plus
		NRL	1	aaggtGatgcc t=86%	minus
		RHOXF1	3	gtgAtgcc t=68%	minus
		RUNX1	1	gatTgtGGcat a=7%	plus
		RUNX3	2	atgCCaCAat t=6%	minus
		SREBF1	1	aTCAccttac a=48%	plus
		SREBF2	2	gTaaggTGAt t=57%	minus
		SREBF2(var.2)	1	gtaAgGTGAt t=47%	minus
		SREBF2(var.2)	1	atCACcTtAc a=73%	plus
		TBX4	1	agGTGatg t=45%	minus
		TBX5	1	agGtGatg t=51%	minus
		TEAD3	6	tgATgCCa T=100%	minus

Table 3 (Continuation)

		TFEC	1	gtaAggtGat t=49%	minus
		THAP1	2	atgCCacaa t=63%	minus
	G	ATF4	1	aggTGAcGccAca c=38%	minus
		ATF7	1	aggTGACGccAcaa C=99%	minus
		ATF7	1	ttgTGgCGTcAcct G=99%	plus
		CEBPβ	1	gTgacgccAc c=0%	minus
		CEBPβ	1	gtggcgtcAc g=80%	plus
		CEBPδ	1	gTgacgccAc c=88%	minus
		CEBPδ	1	gTggcgtcAc g=86%	plus
		CEBPε	1	gTgacgccAc c=81%	minus
		CEBPε	1	gTggcgtcAc g=73%	plus
		CREB1	1	tGAcGcca c=18%	minus
		CREB1	1	tGgcGtca G=91%	plus
		DBP	1	ggTgAcgccAca c=62%	minus
		DBP	1	tgTggcgtcAcc g=55%	plus
		GMEB2	1	tgACGcca C=100%	minus
		HLF	1	ggTgacgccaca c=83%	minus
		HLF	1	tgTggcgtcacc g=67%	plus
		INSM1	1	tgtaaGGtGacg c=8%	minus
		JDP2(var.2)	1	ggTGACGcCAca C=99%	minus
		JDP2(var.2)	1	tgTGgCGTCAcc G=98%	plus

Table 3 (Continuation)

		JUN	1	aaggTGAcGccAc c=16%	minus
		JUN	1	attgTGgcGtcAc G=97%	plus
		JUND (var.2)	1	taaggTGAcgccAca c=31%	minus
		MEIS1	2	gtGACgc C=99%	minus
		MEIS3	1	gtGACgcc C=97%	minus
		MGA	1	tgGcGtcA G=100%	plus
		NFE2L1::MafG	56	ggTGAc c=76%	minus
		NFIX	2	gacGCCAca c=30%	minus
		NR4A2	7	aAGgtgAc c=57%	minus
		RUNX1	1	gatTgtGGcgt g=4%	plus
		RUNX3	1	acgCCaCAat c=9%	minus
		SREBF1	1	gTgAcgccac c=88%	minus
		SREBF1	1	gTCAccttac g=28%	plus
		SREBF2	1	gTaaggTGAc c=34%	minus
		SREBF2	1	gTGgcgTcAc g=77%	plus
		SREBF2(var.2)	1	gtaAgGTGAc c=53%	minus
		SREBF2(var.2)	1	gtCACcTtAc g=27%	plus
		TBX4	4	agGTGacg c=29%	minus
		TBX4	1	tgGcGtca G=100%	plus
		TBX5	4	agGtGacg c=34%	minus
		TFEC	1	gtaAggtGac c=50%	minus
		USF1	1	gtaAggTGacg c=78%	minus

Table 3 (Continuation)

		USF2	1	gtaAgGTGacg c=5%	minus
		ZNF354C	21	cgCCAC c=38%	minus
rs7589621	**G**	BARX1	1	gtacTTAt g=33%	plus
		BSX	1	gtacTTAt g=26%	plus
		DBP	1	aagTAcgTAAag c=62%	minus
		DBP	1	ctTTAcgTActt g=55%	plus
		DLX6	1	gtacTTAt g=26%	plus
		EN1	1	gtAcTtAt g=20%	plus
		EN2	1	cgtacTtAtc g=25%	plus
		ESX1	1	cgtacTtAtc g=7%	plus
		EVX1	1	gataAgTAcg c=27%	minus
		EVX1	1	cgtacTTAtc g=33%	plus
		EVX2	1	gataAgTAcg c=24%	minus
		EVX2	1	cgtacTTAtc g=34%	plus
		FOXA1	1	acttTacgtaCttat g=6%	plus
		GBX1	1	cgtAcTTAtc g=25%	plus
		GBX2	1	cgtAcTTAtc g=31%	plus
		GMEB2	1	gtACGTaa C=98%	minus
		GMEB2	1	ttACGTac G=98%	plus
		GSX1	1	cgtacttAtc g=22%	plus
		HLF	1	aagtacgtaaag c=83%	minus
		HLF	1	ctTtacgtactt g=67%	plus

Table 3 (Continuation)

		HLTF	1	gtaCtTatcc g=20%	plus
		HMBOX1	1	cgtacTtAtc g=15%	plus
		HOXA2	1	cgtacTtAtc g=30%	plus
		HOXB2	1	cgtacTtAtc g=33%	plus
		HOXB3	1	cgtacTTAtc g=31%	plus
		ISL2	1	gtAcTtat g=6%	plus
		ISL2	1	gtAcgtaa c=34%	minus
		ISX	1	gtAcTTAt g=18%	plus
		LBX2	1	cgtacTtAtc g=16%	plus
		MEOX1	1	cgtacTtAtc g=45%	plus
		MEOX2	1	cgtacTtatc g=63%	plus
		MIXL1	1	cgtacTtatc g=20%	plus
		MSX1	1	gtacTtAt g=26%	plus
		NFATC3	1	actTTaCgta g=34%	plus
		NFIL3	1	TTAcGTActta G=91%	plus
		NKX2-3	1	cgtACTTatc g=48%	plus
		NKX2-8	1	gtaCTtatc g=33%	plus
		NKX3-2	1	cgtACTTat g=50%	plus
		NKX6-1	1	gtacTTAt g=33%	plus
		NR3C1	1	aaGtACgtaaaGTgCct C=100%	minus
		NR3C1	1	agGcACtttacGTaCtt G=100%	plus

Table 3 (Continuation)

		NR3C2	1	aaGtACgtaaaGTgCct C=100%	minus
		NR3C2	1	agGcACtttacGTaCtt G=100%	plus
		PDX1	1	gtacTTAt g=34%	plus
		PHOX2A	1	ttAcgtacTta g=20%	plus
		POU2F1	1	agtAcgtaaAgt c=5%	minus
		POU5F1B	1	tAcgtaaAg c=4%	minus
		RORA(var.2)	1	gatAagTacGTaAa c=0%	minus
	A	BARX1	4	atacTTAt a=28%	plus
		BSX	4	atacTTAt a=16%	plus
		DBP	1	aagTAtgTAAag t=38%	minus
		DLX6	4	atacTTAt a=23%	plus
		EN2	1	catacTtAtc a=13%	plus
		ESX1	1	catacTtAtc a=18%	plus
		EVX1	1	gataAgTAtg t=22%	minus
		EVX2	1	gataAgTAtg t=18%	minus
		EVX2	1	catacTTAtc a=25%	plus
		FOXC1	2	aggataAgtAt t=64%	minus
		GMEB2	2	gtAtGTaa t=0%	minus
		GMEB2	2	ttACaTac a=0%	plus
		GSX1	1	catacttAtc a=22%	plus
		HLF	1	ctTtacatactt a=33%	plus
		HLTF	2	ataCtTatcc a=27%	plus

Table 3 (Continuation)

		HNF4G	1	agtatgtAaAGtgCc t=37%	minus
		HMBOX1	1	catacTtAtc a=10%	plus
		HOXA2	1	catacTtAtc a=18%	plus
		HOXB2	1	catacTtAtc a=21%	plus
		HOXB3	1	catacTTAtc a=20%	plus
		ISL2	4	atAcTtat a=28%	plus
		ISX	4	atAcTTAt a=15%	plus
		LBX2	1	catacTtAtc a=12%	plus
		LIN54	2	cTTtacAta A=100%	plus
		NFATC3	1	actTTaCata a=55%	plus
		NFIL3	1	gTAtGTAAagt t=65%	minus
		NEUROD2	3	taCaTActta a=77%	plus
		NKX2-3	1	cacACTTatc a=48%	plus
		NKX2-8	6	ataCTttc a=23%	plus
		NKX3-1	1	catACTTat a=18%	plus
		NKX3-2	1	catACTTat a=30%	plus
		NKX6-1	4	atacTTAt a=18%	plus
		PDX1	4	atacTTAt a=21%	plus
		POU2F1	1	agtATgtaaAgt T=92%	minus
		POU3F1	1	gtATgtaaAgtg T=95%	minus
		POU3F2	1	gtATGtaaAgtg T=95%	minus
		POU3F3	1	agtATGtaaAgtg T=90%	minus
		POU3F4	6	tATGaaAT T=99%	minus

Table 3 (Continuation)

		POU5F1B	2	tATgtaaAg T=92%	minus
		RORA(var.2)	1	gatAagTatGTaAa t=0%	minus
		TBP	1	gtATgtAaagtgcct T=97%	minus
		TEAD1	3	tacATaCtta A=92%	plus
		TEAD3	9	acATaCtt A=100%	plus
		TEAD4	3	tacATaCtta A=94%	plus
rs1868092	**A**	E2F6	2	gaGatGGAggt a=16%	plus
		HLTF	2	ctcCaTctca t=63%	minus
		HLTF	3	gcaCtTtgag a=25%	plus
		HNF4G	1	ccatctCAaAGtgca t=30%	minus
		HOXA5	10	ctcaaagt t=25%	minus
		NKX2-3	1	tgCACTTtga a=34%	plus
		NR2C2	1	ccatctcaaaGtgca t=81%	minus
		NKX2-8	3	gcaCTttga a=37%	plus
		SOX10	91	cttTga a=0%	plus
		THAP1	5	cctCCatct t=16%	minus
		YY1	1	tgAgATGGaggt A=94%	plus
		YY2	1	gacCtCCATct t=40%	minus
	G	E2F6	1	ggGatGGAggt g=69%	plus
		HIC2	1	aTcCCacaa C=99%	minus
		HLTF	5	gcaCtTtggg g=25%	plus
		HNF4G	1	ccatccCAaAGtgca c=30%	minus
		KLF5	1	ctccatCCCa C=100%	minus

Table 3 (Continuation)

		KLF5	2	cctcCatCCc C=97%	minus
		MZF1	85	ttGGGA G=95%	plus
		NFIA	1	caTcCCAAag C=100%	minus
		NFIC	85	tTGGga G=96%	plus
		NFIX	2	catcCCAaa C=90%	minus
		NKX2-3	1	tgCACTTtgg g=30%	plus
		NKX2-8	7	gcaCTttgg g=30%	plus
		SOX10	76	cttTgg g=5%	plus
		TEAD1	1	tccATcCcaa C=95%	minus
		THAP1	2	catCCcaaa C=98%	minus
		THAP1	3	cctCCatcc c=17%	minus

regulation of carbohydrate, protein and fat metabolism, mediates aldosterone actions on salt and water balance, and catecholamine biosynthetic genes, respectively (Tables 2 & 3). The minor rs7589621 SNP *EPAS1*-A allele creates eleven unique punitive TFBS for the FOXC1, HNF4G, LIN54, NEUROD2, POU3F1-4, TEAD1, and TEAD3 & 4, TFs which are involved with cell viability and resistance to oxidative stress, transcriptional repression, regulation of cell cycle genes, neuronal determination, early embryogenesis and neurogenesis, and enhancer for transcription, respectively (Tables 2 & 3). There are also thirty-one conserved TFBS for the BARX1, BSX, DBP, DLX6, EN2, ESX1, EVX1 & 2, GMEB2, GSX1, HLF, HLTF, HMBOX1, HOXA2, HOXB2, HOXB3, ISL2, ISX, LBX2, NFATC3, NKX2-3, RORA AND TBP TFs which are involved with craniofacial development, transcriptional activation, circadian rhythm, roles in forebrain, central nervous system, placental development, specification of neuronal cell types, DNA replication, pituitary development, transcription activation, altering chromatin structure, transcriptional repressor, regulates development, axon pathways, regulates gene expression in the intestine, induces expression of cytokine genes in T-cells, homeodomain, nuclear hormone receptors, and the pre-initiation complex, respectively (Tables 2 & 3).

The common rs1868092 SNP *EPAS1*-A allele creates four unique TFBS for the HOXA5, NR2C2, YY1 & 2 TFs which are involved with the development regulatory system, repression or activation of transcription, and positive and negative control of transcription at the transcription start site, respectively (Tables 2 & 3). The minor *EPAS1*-G allele creates seven unique TFBS for the HIC2, KLF5, MZF1, NFIA, NFIC, NFIX and TEAD1 TFs which are involved with transcriptional repression, transcription, hemopoietic development, transcription and replication, and enhancer of transcription, respectively (Tables 2 & 3). There are also seven conserved punitive TFBS for the E2F6, HLTF, HNF4G, NKX2-3, NKX2-8, SOX10 and THAP1 TFs which are involved in tumor suppression, chromatin structure, transcriptional activation, homeodomain, regulatory, and regulates endothelial cell proliferation, respectively (Tables 2 & 3).

Discussion:

Genome-wide association studies (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 [44, 45] and the remaining 93% are located within non-coding areas [46, 47] 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 [23, 25, 48]. 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 [48]. Examples of rSNPs associated with disease susceptibility are numerous and several reviews have been published [48-51].

The rs56721780 rSNP *EPAS1*-G allele [G (+ strand) or C (- strand)] located in the unique RELA, RUNX1, TFAP2A, B & C TFBS have a 100%, 92% and 94 -100% occurrence, respectively in humans (Table 3). Since these binding sites (BS) occur only once in the gene except for the RUNX1 TFBS which occurs twice, the rSNP G allele should have a tremendous impact on gene regulation by these TFs (Table 3). The minor rs56721780 rSNP *EPAS1*-C allele [C (+ strand) or G (-

strand)] located in the unique FOXP3 and HOXA5 TFBS have a 65% and 31% occurrence, respectively in humans. Since these TFBS have a low occurrence in humans and occur more than once in the gene, the respective TF would not be expected to have much impact on the regulation of the *EPAS1* gene (Figure 1, Table 3).

The rs6756667 rSNP *EPAS1*-A allele [A (+ strand) or T (- strand) located in the unique CEBPa, FOS::JUN, HIC 1 & 2, HOXA2, NFIA and NRL TFBS has a 78, 94, 95, 98, 90, 100, 86 and 100% occurrence, respectively in humans (Figure 2, Table 3); however, only the CEBPa, NFIA and NRL TFBS occur only once in the gene. Consequently the corresponding three TFs should impact the regulation of the *EPAS1* gene. The minor rs6756667 rSNP *EPAS1*-G allele [G (+ strand) or T (- strand)] located in the unique ATF7, GMEB2, JDP2, and MGA TFBS have a 99, 100, 99, 100% occurrence, respectively in humans and all only occur once in the *EPAS1* gene. Consequently, the TFs for these TFBS could have some impact on the regulation of the *EPAS1* gene (Tables 2 & 3).

The rs7589621 rSNP *EPAS1*-G allele [G (+ strand) or C (- strand)] located in the unique NR3C1 & 2 TFBSs have a 100% occurrence in humans and are found only once in the *EPAS1* gene (Table 3). Consequently, the corresponding glucocorticoid and mineralocorticoid nuclear receptors which bind their respective BS should have a major impact on the regulation of the EPAS1 gene (Tables 2 & 3). The minor rs7589621 rSNP *EPAS1*-A allele [A (+ strand) or T (- strand)] located in the unique LIN54, POU3F1-4, TEAD1, 3, 4 TFBS has a 100, 95, 95, 90, 99, 92, 100 and 94% occurrence, respectively in humans (Table 3). However, only the POU3F1-3 TFBS occur once in the EPAS1 gene, consequently, their corresponding TFs should have an impact on the *EPAS1* gene regulation. The remaining TFBS occur multiple times in the gene and would not be expected to have much impact on gene regulation (Figure 2, Tables 2 & 3).

The rs1868092 rSNP *EPAS1*-A allele [A (+ strand) or T (- strand)] located in the unique NR2C2 and YY1 TFBS have a 81 and 94%% occurrence, respectively in humans and only occur once in the *EPAS1* gene (Tables 2 & 3). The NR2C2 orphan nuclear receptor which can occur as a repressor of activator of transcription and the YY1 TF which exhibits both positive and negative control of transcript should have an impact on the regulation of the *EPAS1* gene (Table 2). The minor rs1868092 rSNP *EPAS1*-G allele [G (+ strand) or C (- strand)] located in the unique HIC2, KLF5, NFIA, and TEAD1 TFBS have a 99, 100, 100 and 95% occurrence, respectively in humans and occur only once in the *EPAS1* gene (Table 3). Since the corresponding TFs function as activators, enhancers and repressors, the occurrence of these TFBS should impact the regulation of the *EPAS1* gene (Tables 2 & 3).

Human diseases or conditions can be associated with rSNPs of the *EPAS1* gene as illustrated above. What a change in the rSNP alleles can do, is to alter the DNA landscape around the SNP for potential TFs to attach and regulate a gene. As an example, the punitive TFBS associated with the rs56721780 common rSNP EPAS1-G allele from Table 3 as illustrated in Figure 1 as well as the rs7589621 common rSNP *STAT4*-G allele as illustrated in Figure 2. As can be seen in Table 3, these potential TFBS change when an individual carries the alternate allele. The importance of this has been illustrated in Figure 1 with the punitive TFAP2A, B & C TFBS where the common G allele has binding sites for these TFs and the minor C allele does not. The TFAP2A, B & C TFs act as activators and repressors and are involved in a large spectrum of biological functions such as proper eye, face, body wall, limb and neural tube development (cf. Table 2). Another example would be the punitive NR3C1 & 2 TFBS where the common rs7589621 rSNP G allele has created these binding sites

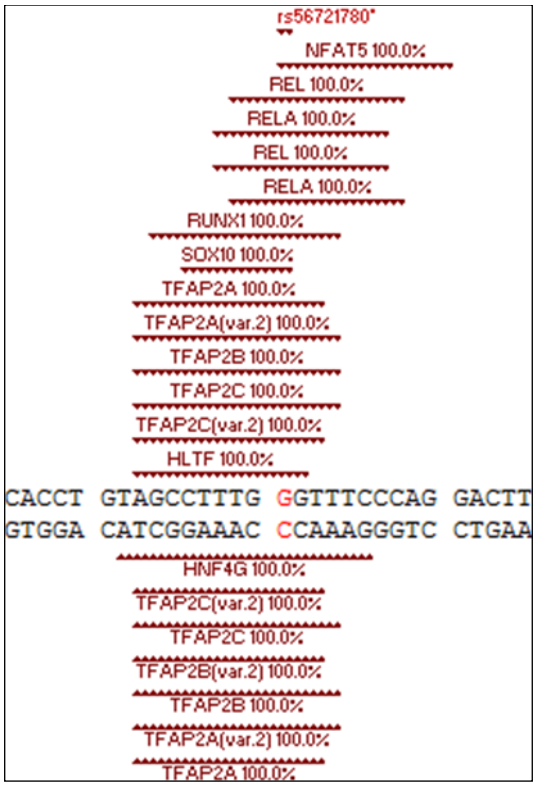

Figure 1. Double stranded DNA from the *EPAS1* gene showing the potential TFBS for twenty different TFs which can bind their respective DNA sequence either above (+) or below (-) the duplex (cf. Table 3). The rs56721780 rSNP common *EPAS1*-G allele is found in each of these TFBS. As shown, this rSNP is located in the promoter region of the *EPAS1* gene. Also included with the potential TFBS is their % sequence homology to the duplex.

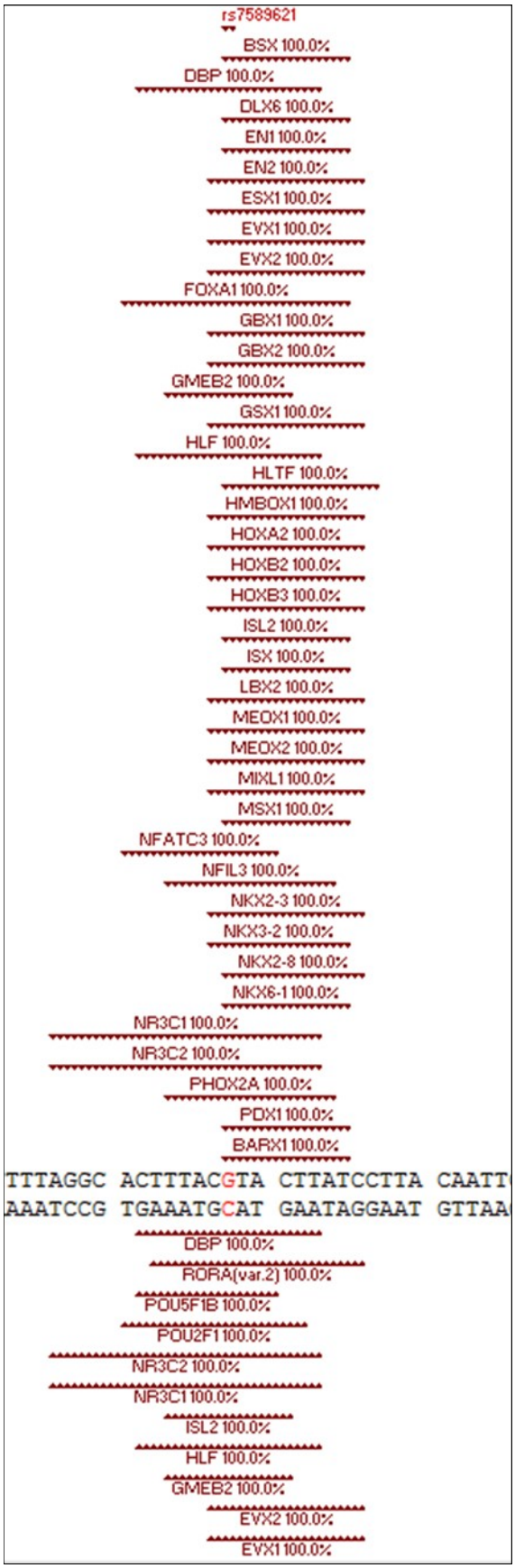

Figure 2. Double stranded DNA from the *EPAS1* gene showing the potential TFBS for forty eight different TFs which can bind their respective DNA sequence either above (+) or below (-) the duplex (cf. Table 3). The rs7589621 rSNP common *EPAS1*-G allele is found in each of these TFBS. As shown, this rSNP is located in intron two of the *EPAS1* gene. Also included with the potential TFBS is their % sequence homology to the duplex.

for the glucocorticoid and mineralocorticoid nuclear receptors which regulate carbohydrate, protein and fat metabolism while the minor A allele has eliminated these TFBS (Figure 2, Tables 2 & 3). Other examples can be found in Table 3.

Conclusion:

SNPs that alter the TFBS are not only found in the promoter regions but in the introns, exons and the UTRs of a gene. The nucleus of the cell is where epigenetic alterations occur 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 disease. In this report, examples have been described to illustrate that a change in rSNP alleles in the *EPAS1* gene can provide different TFBS which in turn are also associated with human disease or alterations in human health such as adaptions to high altitude. The punitive changes in TFBS created by the four rSNPs could very well influence the significant cline in allele frequencies seen in Tibetans with increasing altitude [20] or the haplotype association with high altitude polycythemia in male Han Chinese [22]. As an example, the minor rs7589621 SNP *EPAS1*-A creates a potential TFBS for the FOXC TF which is an important regulator of cell viability and resistance to oxidative stress. Where oxidative stress is linked to oxygen, hypoxia, heart failure and the hypoxia-inducible factor transcriptional factors [52]. The potential alterations in TFBS obtained by computational analyses need to be verified by future protein/DNA electrophoretic mobility gel shift assays and gene expression studies.

Competing Interest

Author has declared that no competing interests exist.

References:

1. Ward MP, Everest 1953, first ascent: a clinical record. High Alt Med Biol 2003;4: 27-37.
2. Loboda A, Jozkowicz A, Dulak J, HIF-1 versus HIF-2--is one more important than the other? Vascul Pharmacol;56: 245-51.
3. Giaccia AJ, Simon MC, Johnson R, The biology of hypoxia: the role of oxygen sensing in development, normal function, and disease. Genes Dev 2004;18: 2183-94.
4. Semenza GL, HIF-1 and mechanisms of hypoxia sensing. Curr Opin Cell Biol 2001;13: 167-71.
5. Schofield CJ, Ratcliffe PJ, Oxygen sensing by HIF hydroxylases. Nat Rev Mol Cell Biol 2004;5: 343-54.
6. Peng Y, Yang Z, Zhang H, Cui C, Qi X, Luo X, Tao X, Wu T, Ouzhuluobu, Basang, Ciwangsangbu, Danzengduojie, Chen H, Shi H, Su B, Genetic variations in Tibetan populations and high-altitude adaptation at the Himalayas. Mol Biol Evol;28: 1075-81.
7. Xu S, Li S, Yang Y, Tan J, Lou H, Jin W, Yang L, Pan X, Wang J, Shen Y, Wu B, Wang H, Jin L, A genome-wide search for signals of high-altitude adaptation in Tibetans. Mol Biol Evol;28: 1003-11.
8. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, Pool JE, Xu X, Jiang H, Vinckenbosch N, Korneliussen TS, Zheng H, Liu T, He W, Li K, Luo R, Nie X, Wu H, Zhao M, Cao H, Zou J, Shan Y, Li S, Yang Q, Asan, Ni P, Tian G, Xu J, Liu X, Jiang T, Wu R, Zhou G, Tang M, Qin J, Wang T, Feng S, Li G, Huasang, Luosang J, Wang W, Chen F, Wang Y, Zheng X, Li Z, Bianba Z, Yang G, Wang X, Tang S, Gao G, Chen Y, Luo Z,

Gusang L, Cao Z, Zhang Q, Ouyang W, Ren X, Liang H, Huang Y, Li J, Bolund L, Kristiansen K, Li Y, Zhang Y, Zhang X, Li R, Yang H, Nielsen R, Wang J, Sequencing of 50 human exomes reveals adaptation to high altitude. Science;329: 75-8.

9. Xu XH, Huang XW, Qun L, Li YN, Wang Y, Liu C, Ma Y, Liu QM, Sun K, Qian F, Jin L, Wang J, Two functional loci in the promoter of EPAS1 gene involved in high-altitude adaptation of Tibetans. Sci Rep;4: 7465.

10. Semenza GL, Life with oxygen. Science 2007;318: 62-4.

11. Majmundar AJ, Wong WJ, Simon MC, Hypoxia-inducible factors and the response to hypoxic stress. Mol Cell;40: 294-309.

12. Skuli N, Simon MC, HIF-1alpha versus HIF-2alpha in endothelial cells and vascular functions: is there a master in angiogenesis regulation? Cell Cycle 2009;8: 3252-3.

13. Skuli N, Liu L, Runge A, Wang T, Yuan L, Patel S, Iruela-Arispe L, Simon MC, Keith B, Endothelial deletion of hypoxia-inducible factor-2alpha (HIF-2alpha) alters vascular function and tumor angiogenesis. Blood 2009;114: 469-77.

14. Semenza GL, Oxygen sensing, homeostasis, and disease. N Engl J Med;365: 537-47.

15. Eltzschig HK, Carmeliet P, Hypoxia and inflammation. N Engl J Med;364: 656-65.

16. Kaelin WG, Jr., The von Hippel-Lindau tumour suppressor protein: O2 sensing and cancer. Nat Rev Cancer 2008;8: 865-73.

17. Han SS, Yeager M, Moore LE, Wei MH, Pfeiffer R, Toure O, Purdue MP, Johansson M, Scelo G, Chung CC, Gaborieau V, Zaridze D, Schwartz K, Szeszenia-Dabrowska N, Davis F, Bencko V, Colt JS, Janout V, Matveev V, Foretova L, Mates D, Navratilova M, Boffetta P, Berg CD, Grubb RL, 3rd, Stevens VL, Thun MJ, Diver WR, Gapstur SM, Albanes D, Weinstein SJ, Virtamo J, Burdett L, Brisuda A, McKay JD, Fraumeni JF, Jr., Chatterjee N, Rosenberg PS, Rothman N, Brennan P, Chow WH, Tucker MA, Chanock SJ, Toro JR, The chromosome 2p21 region harbors a complex genetic architecture for association with risk for renal cell carcinoma. Hum Mol Genet;21: 1190-200.

18. Xue X, Taylor M, Anderson E, Hao C, Qu A, Greenson JK, Zimmermann EM, Gonzalez FJ, Shah YM, Hypoxia-inducible factor-2alpha activation promotes colorectal cancer progression by dysregulating iron homeostasis. Cancer Res;72: 2285-93.

19. Tian H, McKnight SL, Russell DW, Endothelial PAS domain protein 1 (EPAS1), a transcription factor selectively expressed in endothelial cells. Genes Dev 1997;11: 72-82.

20. Basang Z, Wang B, Li L, Yang L, Liu L, Cui C, Lanzi G, Yuzhen N, Duo J, Zheng H, Wang Y, Xu S, Jin L, Wang X, HIF2A Variants Were Associated with Different Levels of High-Altitude Hypoxia among Native Tibetans. PLoS One;10: e0137956.

21. Guo LI, Zhang J, Jin J, Gao X, Yu J, Geng Q, Li H, Huang L, Genetic variants of endothelial PAS domain protein 1 are associated with susceptibility to acute mountain sickness in individuals unaccustomed to high altitude: A nested case-control study. Exp Ther Med;10: 907-14.

22. Chen Y, Jiang C, Luo Y, Liu F, Gao Y, An EPAS1 haplotype is associated with high altitude polycythemia in male Han Chinese at the Qinghai-Tibetan plateau. Wilderness Environ Med;25: 392-400.

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

24. Knight JC, Regulatory polymorphisms underlying complex disease traits. Journal of molecular medicine 2005;83: 97-109.

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

26. Wang X, Tomso DJ, Chorley BN, Cho HY, Cheung VG, Kleeberger SR, Bell DA, Identification of polymorphic antioxidant response elements in the human genome. Hum Mol Genet 2007;16: 1188-200.

27. Claessens F, Verrijdt G, Schoenmakers E, Haelens A, Peeters B, Verhoeven G, Rombauts W, Selective DNA binding by the androgen receptor as a mechanism for hormone-specific gene regulation. The Journal of steroid biochemistry and molecular biology 2001;76: 23 -30.

28. Hsu MH, Savas U, Griffin KJ, Johnson EF, Regulation of human cytochrome P450 4F2 expression by sterol regulatory element-binding protein and lovastatin. J Biol Chem 2007;282: 5225-36.

29. Takai H, Araki S, Mezawa M, Kim DS, Li X, Yang L, Li Z, Wang Z, Nakayama Y, Ogata Y, AP1 binding site is another target of FGF2 regulation of bone sialoprotein gene transcription. Gene 2008;410: 97-104.

30. Buroker NE, Huang JY, Barboza J, Ledee DR, Eastman RJ, Jr., Reinecke H, Ning XH, Bassuk JA, Portman MA, The adaptor-related protein complex 2, alpha 2 subunit (AP2alpha2) gene is a peroxisome proliferator-activated receptor cardiac target gene. The protein journal 2012;31: 75-83.

31. Huang CN, Huang SP, Pao JB, Hour TC, Chang TY, Lan YH, Lu TL, Lee HZ, Juang SH, Wu PP, Huang CY, Hsieh CJ, Bao BY, Genetic polymorphisms in oestrogen receptor-binding sites affect clinical outcomes in patients with prostate cancer receiving androgen-deprivation therapy. Journal of internal medicine 2012;271: 499-509.

32. Huang CN, Huang SP, Pao JB, Chang TY, Lan YH, Lu TL, Lee HZ, Juang SH, Wu PP, Pu YS, Hsieh CJ, Bao BY, Genetic polymorphisms in androgen receptor-binding sites predict survival in prostate cancer patients receiving androgen-deprivation therapy. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO 2012;23: 707-13.

33. Yu B, Lin H, Yang L, Chen K, Luo H, Liu J, Gao X, Xia X, Huang Z, Genetic variation in the Nrf2 promoter associates with defective spermatogenesis in humans. Journal of molecular medicine 2012.

34. Wu J, Richards MH, Huang J, Al-Harthi L, Xu X, Lin R, Xie F, Gibson AW, Edberg JC, Kimberly RP, Human FasL gene is a target of beta-catenin/T-cell factor pathway and complex FasL haplotypes alter promoter functions. PLoS One 2011;6: e26143.

35. Alam M, Pravica V, Fryer AA, Hawkins CP, Hutchinson IV, Novel polymorphism in the promoter region of the human nerve growth-factor gene. International journal of immunogenetics 2005;32: 379-82.

36. Kumar A, Purohit R, Computational investigation of pathogenic nsSNPs in CEP63 protein. Gene 2012;503: 75-82.

37. Kamaraj B, Purohit R, Computational screening of disease-associated mutations in OCA2 gene. Cell Biochem Biophys 2014;68: 97-109.

38. Kumar A, Rajendran V, Sethumadhavan R, Shukla P, Tiwari S, Purohit R, Computational SNP analysis: current approaches and future prospects. Cell Biochem Biophys 2014;68: 233-9.

39. Kumar A, Purohit R, Use of long term molecular dynamics simulation in predicting cancer associated SNPs. PLoS Comput Biol 2014;10: e1003318.

40. Bryne JC, Valen E, Tang MH, Marstrand T, Winther O, da Piedade I, Krogh A, Lenhard B, Sandelin A, JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. Nucleic Acids Res 2008;36: D102-6.

41. Sandelin A, Alkema W, Engstrom P, Wasserman WW, Lenhard B, JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res 2004;32: D91 -4.

42. Sandelin A, Wasserman WW, Lenhard B, ConSite: web-based prediction of regulatory elements using cross-species comparison. Nucleic Acids Res 2004;32: W249-52.

43. Buroker NE, Ning XH, Zhou ZN, Li K, Cen WJ, Wu XF, Zhu WZ, Scott CR, Chen SH, AKT3, ANGPTL4, eNOS3, and VEGFA associations with high altitude sickness in Han and Tibetan Chinese at the Qinghai-Tibetan Plateau. International journal of hematology;96: 200-13.

44. Pennisi E, The Biology of Genomes. Disease risk links to gene regulation. Science 2011;332: 1031.

45. Kumar V, Wijmenga C, Withoff S, From genome-wide association studies to disease mechanisms: celiac disease as a model for autoimmune diseases. Semin Immunopathol 2012;34: 567-80.

46. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA, Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A 2009;106: 9362-7.

47. Kumar V, Westra HJ, Karjalainen J, Zhernakova DV, Esko T, Hrdlickova B, Almeida R, Zhernakova A, Reinmaa E, Vosa U, Hofker MH, Fehrmann RS, Fu J, Withoff S, Metspalu A, Franke L, Wijmenga C, Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. PLoS Genet 2013;9: e1003201.

48. Chorley BN, Wang X, Campbell MR, Pittman GS, Noureddine MA, Bell DA, Discovery and verification of functional single nucleotide polymorphisms in regulatory genomic regions: current and developing technologies. Mutat Res 2008;659: 147-57.

49. Prokunina L, Alarcon-Riquelme ME, Regulatory SNPs in complex diseases: their identification and functional validation. Expert Rev Mol Med 2004;6: 1-15.

50. Buckland PR, The importance and identification of regulatory polymorphisms and their mechanisms of action. Biochim Biophys Acta 2006;1762: 17-28.

51. Sadee W, Wang D, Papp AC, Pinsonneault JK, Smith RM, Moyer RA, Johnson AD, Pharmacogenomics of the RNA world: structural RNA polymorphisms in drug therapy. Clin Pharmacol Ther 2011;89: 355-65.

52. Giordano FJ, Oxygen, oxidative stress, hypoxia, and heart failure. J Clin Invest 2005;115: 500-8.