

Identifying Alterations in Transcriptional Factor Binding Sites

Buroker NE*

Department of Pediatrics, University of Washington, USA

Abstract

SNPs located in non-coding regions of the human genome that have been found to be significantly associated with disease or sickness may affect gene regulation by altering binding sites for Transcription Factors (TF). There have been several SNP studies conducted to identify punitive changes in Transcriptional Factor Binding Sites (TFBS) that may affect gene regulation but conformation of these changes has been lacking. In conjunction with these previous SNP studies, the HaploReg and RegulomeDB databases were explored for conformational TF motif and protein binding changes that would affect gene expression. This investigation has led to a conformation between some punitive TFBS changes from the previous studies and experimental evidence found in these databases verifying changes. These changes are discussed with relation to alterations in gene expression that may result in disease and sickness.

Introduction

Non-coding DNA comprises 98% of the human genome [1]. Genome-Wide Association Studies (GWAS) indicate that approximately 93% of the disease or trait-predisposing Single Nucleotide Polymorphisms (SNPs) believed to be associated with gene regulation fall in these non-coding regions [2-4]. SNPs that have been found to be statistically significant for a disease or sickness in a population are considered to be risk-associated SNPs that cause changes in gene expression levels [4]. A single nucleotide change in a transcriptional factor motif sequence or a Transcriptional Factor Binding Site (TFBS) may affect the process of gene regulation [5-7]. A SNP occurring in a motif may increase or decrease the corresponding Transcription Factor's (TF) ability to bind DNA resulting in allele-specific gene expression [8]. Several studies have been made to identify changes in punitive TFBS created by regulatory (r) SNPs significantly associated with human disease or sickness [9-20]. However, among these studies to date, there hasn't been any conclusive evidence as to which TF alteration might be responsible for a given disease or sickness. In an effort to resolve this issue in the previous studies, the HaploReg [21] and RegulomeDB [22] databases were screened for the TFBS created by a causative allele substitution that could alter gene regulation resulting in disease or sickness.

The HaploReg database is a tool for examining annotations of the non-coding genome among published GWAS studies while the RegulomeDB guides interpretation of regulatory variants in the human genome using experimental data sets from the ENCODE project. RegulomeDB also makes computational predictions to identify putative regulatory potential and functional variants. In present report, these analyses were conducted on nine previously studied genes whose rSNPs have significantly associated with disease or sickness [9-20].

Materials and methods

Genes and rSNPs

Nine genes (see Results) whose rSNPs have been shown to be significantly associated with disease or sickness [9-20] are listed in

the Table as well as the rSNP alleles and frequencies. The number of conserved TFBS between the two SNP alleles and unique TFBS per allele which have been previously determined [9-20] are included in the Table. The names of conserved and unique TFBS can be found in the previous studies [9-20] as referenced below the gene symbol in the Table.

Identifying TFBS

The JASPAR CORE database [23, 24] and ConSite [25] were used to identify the TFBS in previous studies [9-20]. JASPAR is a collection of transcription factor DNA-binding preferences used for scanning genomic sequences where ConSite is a web-based tool for finding cis-regulatory elements in genomic sequences. The Vector NTI Advance 11 computer program (Invitrogen, Life Technologies) was used to locate SNPs and TFBS within all genes listed in the Table.

HaploReg and RegulomeDB

HaploReg in part is a resource for exploring regulatory motif alterations within sets of genetically linked variants [21]. RegulomeDB guides interpretation of regulatory variants in the human genome [22]. The two databases were used to examine the changes created by the SNPs within the regulatory regions of the genes listed in the Table. A conformation of the TFBS from the two databases with that generated from the previous studies [9-20] are reported in the Table. See the Supplement for transcription factor and/or protein descriptions or function.

Results

The adrenergic, beta, receptor kinase 1 (*ADRBK1*) gene an important regulator of adrenergic signaling which plays a central role in

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

Key words: gene regulation; non-coding DNA; rSNPs; TFBS; human disease

Received: February 14, 2017; **Accepted:** February 21, 2017; **Published:** February 24, 2017

Table: Nine genes whose rSNPs have previously been shown to be associated with disease or sickness. Listed are the SNP alleles, the number of conserved punitive TFBS for the alleles and unique TFBS per each allele. See references for TFBS name, description and DNA sequence. Also listed are the TF motif changes and proteins bound that match the punitive TFBS per SNP allele which were obtained from the HaploReg and RegulomeDB databases of experimental data. MAF is minor allele frequency. See supplement for protein/transcriptional factor name and description or function.

Gene (Ref)	rSNP	Alleles (MAF)	Conserved TFBS	Unique TFBS	Motif Changes		Proteins bound HaploReg & RegulomeDB
					HaploReg	RegulomeDB	
ADRBK1 11	rs948988	G	7	2			
		A(0.29)		10	NFE2		
	rs4370946	C	3	12	NFKB1		
AKT3 9		T(0.2)		5			
	rs4590656	C	2	2	Foxa		
		T(0.40)		4	Foxa		
	rs12031994	G	6	8	IRF, Sox		
		A(0.17)		8	IRF, Sox		
	rs10157763	C	4	5			
ATF3 10		T(0.4)		4	CTCF	CTCF	
	rs2125230	G	2	4			
		A(0.28)		9	Foxa		
	rs11119982	C	6	1			
		T(0.32)		5	NKX2		
DIO2 12	rs12885300	C	4	0	PRDM1		
		T(0.23)		6	PRDM1		
	rs225010	C	5	6	Pax		
		T(0.42)		4	Pax		
	rs225011	C	8	2	RORa1	RORa1	
		T(0.42)		3	RORa1	RORa1	
	rs225013	G	1	2			
EPAS1 13		T(0.36)		9	HNF1, Pax		
	rs225014	T	4	5	RXRa		CTCF,RAD21,SMC3,MAX,YY1,cMYC
		C(0.42)		6	RXRa		CTCF,RAD21,SMC3,MAX,YY1,cMYC
	rs6574549	T	7	5	Arid3a,Nkx2,Nkx3	Nkx2,Nkx3	
		G(0.01)		4	Foxa,Nkx2,Nkx3, Pou2F2	Nkx2,Nkx3	
	rs6756667	A	22	13	ATF4		
LIPA 14		G(0.31)		9	ATF4		
	rs7589621	G	27	10	Nkx2, Nkx3, Pou2f2	GMEB2, Nkx2-3	
		A(0.25)		11	Nkx2, Nkx3, Pou2f2	GMEB2, Nkx2-3	
STAT4 15	rs2246833	C	5	3	ZNF143	ZNF143	CTCF
		T(0.42)		1			CTCF
	rs1412444	C	13	5	Ets1, FEV	ELK1,ELK4,Ets1, FEV,FLI1,GABPA	
		T(0.42)		12	FEV	ELK1,ELK4,FEV,FLI1,GABPA	
TBXA2R 26	rs11889341	C	7	3			
		T(0.34)		15	Foxo1		
	rs8179673	T	8	9	Arid3a,Sox,TBP	Fox, Sox, Tbp	
		C(0.26)		15	HNF1, HNF4, Sox, TBP	Fox, Sox, Tbp	
	rs7572482	A	4	8	Pou5f1,Sox2	Pou5f1	EBF1
VEGFA 16,18-20		G(0.47)		4			
	rs2238633	G	2	2	Klf4		
		T(0.22)		2	Klf4		
	rs2238634	G	1	0	ZFX		
VEGFA 16,18-20		T(0.22)		5	HNF4, ZFX		HNF4
	rs4523	T	5	4	Ets1, GR		
		C(0.20)		5	Ets1		
VEGFA 16,18-20	rs699947	C	2	3		Tcfcp211	
		A(0.17)		1		Tcfcp211	
	rs79469752	C	1	3	Pax5		Pax5
		T(0.05)		6	Pax5		Pax5
	rs13207351	G	1	1	Nrf1		Nrf1
		A(0.31)		2	Nrf1	Pax5	Nrf1, Pax5
	rs28357093	A	1	2	Nrf1	E2F6	E2F6
		C(0.1)		2	Nrf1		
VEGFA 16,18-20	rs1570360	G	5	3			EGR1, SP1
		A(0.09)		6			SP1
	rs2010963	C	1	3	IRF1		
	G(0.43)		1				

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.
ATF4	Activating Transcription Factor 4	Transcriptional activator. Binds the cAMP response element (CRE) (consensus: 5-GTGACGT[AC][AG]-3), a sequence present in many viral and cellular promoters.
CTCF	CCCTC-Binding Factor	Chromatin binding factor that binds to DNA sequence specific sites. Involved in transcriptional regulation by binding to chromatin insulators and preventing interaction between promoter and nearby enhancers and silencers. The protein can bind a histone acetyltransferase (HAT)-containing complex and function as a transcriptional activator or bind a histone deacetylase (HDAC)-containing complex and function as a transcriptional repressor.
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 evolutionarily 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 differentiation and mitogenesis.
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 serum response factor and the serum response element in the promoter of the c-fos proto-oncogene.
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.
FLI1	Fli-1 Proto-Oncogene, ETS Transcription Factor	This gene encodes a transcription factor containing an ETS DNA-binding domain.
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.
FOXO1	Forkhead Box O1	Transcription factor that is the main target of insulin signaling and regulates metabolic homeostasis in response to oxidative stress.
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 Transcription Factor Alpha Subunit	This gene encodes one of three GA-binding protein transcription factor subunits which functions as a DNA-binding subunit. Since this subunit shares identity with a subunit encoding the nuclear respiratory factor 2 gene, it is likely involved in activation of cytochrome oxidase expression and nuclear control of mitochondrial function.
GMEB2	Glucocorticoid Modulatory Element Binding Protein 2	Modulation of transactivation by the glucocorticoid receptor bound to glucocorticoid response elements.
GR	Glucocorticoid receptor Nuclear Receptor Subfamily 3 Group C Member 1	This gene encodes glucocorticoid receptor, which can function both as a transcription factor that binds to glucocorticoid response elements in the promoters of glucocorticoid responsive genes to activate their transcription, and as a regulator of other transcription factors.
HNF1A	Hepatocyte nuclear factor 1, alpha	Transcriptional activator that regulates the tissue specific expression of multiple genes, especially in pancreatic islet cells and in liver.
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.
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).

KLF4	Krüppel-like factor 4	Transcription factor that can act both as activator and as repressor. Regulates the expression of key transcription factors during embryonic development.
MAX	MYC Associated Factor X	Transcription regulator. Forms a sequence-specific DNA-binding protein complex with MYC or MAD which recognizes the core sequence 5-CAC[GA]TG-3.
MYC	V-Myc Avian Myelocytomatosis Viral Oncogene Homolog	The protein encoded by this gene is a multifunctional, nuclear phosphoprotein that plays a role in cell cycle progression, apoptosis and cellular transformation.
MYC::MAX	v-myc avian myelocytomatosis viral 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.
NFE2	Nuclear Factor, Erythroid 2	Component of the NF-E2 complex essential for regulating erythroid and megakaryocytic maturation and differentiation.
NFKB1	Nuclear factor of kappa light polypeptide oncogene homolog 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.
NKX2-1	NK2 Homeobox 1	Transcription factor that binds and activates the promoter of thyroid specific genes such as thyroglobulin, thyroperoxidase, and thyrotropin receptor.
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.
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.
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.
PAX1	Paired box gene 1	This protein is a transcriptional activator. It may play a role in the formation of segmented structures of the embryo.
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.
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'-ATTGTCAT-3', a common transcription factor binding site in immunoglobulin gene promoters.
POU5F1	POU Class 5 Homeobox 1	Transcription factor that binds to the octamer motif (5-ATTGTCAT-3). 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.
PRDM1	PR domain containing 1, with ZNF domain	This gene encodes a protein that acts as a repressor of beta-interferon gene expression.
RAD21	RAD21 Cohesin Complex Component	Cleavable component of the cohesin complex, involved in chromosome cohesion during cell cycle, in DNA repair, and in apoptosis.
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 skeletal muscle development.
RXR α	Retinoid X Receptor Alpha	Receptor for retinoic acid. Retinoic acid receptors bind as heterodimers to their target response elements in response to their ligands, all-trans or 9-cis retinoic acid, and regulate gene expression in various biological processes.
SMC3	Structural Maintenance of Chromosomes 3	This gene belongs to the SMC3 subfamily of SMC proteins. The encoded protein occurs in certain cell types as either an intracellular, nuclear protein or a secreted protein. Central component of cohesin, a complex required for chromosome cohesion during the cell cycle.
SOX1	SRY (sex determining region Y)-box 1	This intronless gene encodes a member of the SRY-related HMG-box (SOX) family of transcription factors involved in the regulation of embryonic development and in the determination of cell fate.
SOX2	SRY (sex determining region Y)-box 2	Transcription factor that forms a trimeric complex with OCT4 on DNA and controls the expression of a number of genes involved in embryonic development such as YES1, FGF4, UTF1 and ZFP206 (By similarity).
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.
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.
TCF21	Transcription factor CP2-like 1	Transcriptional suppressor. May suppress UBP1-mediated transcriptional activation. Modulates the placental expression of CYP11A1.
YY1	YY1 transcription factor	YY1 is a ubiquitously distributed transcription factor belonging to the GLI-Krüppel 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.
ZFX	Zinc finger X-chromosomal protein	A member of the krüppel C2H2-type zinc-finger protein family and probable transcriptional activator.
ZNF143	Zinc finger protein 143	Transcriptional activator. Activates the gene for selenocysteine tRNA (tRNA ^{Sec}). Interaction with CHD8. Binds to the SPH motif of small nuclear RNA (snRNA) gene promoters. Participates in efficient U6 RNA polymerase III transcription via its interaction with CHD8.

heart failure pathology, has two (rs948988 and rs4370947) rSNPs whose punitive TFBS [11] were analyzed for binding site verification with the HaploReg and RegulomeDB databases (Table). The two alleles (G/A) of the rs948988 rSNP generates seven conserved punitive TFBS between the two alleles while the common G allele generates an additional two unique TFBS and the minor A allele generates an additional 10 unique TFBS. Of the 10-punitive unique TFBS generated by the A allele, the nuclear factor erythroid 2-related factor (NFE2) BS is also found in the HaploReg database for this rSNP (Table). NFE2 coordinates the up-regulation of cytoprotective genes via the antioxidant response element (ARE). The two alleles (C/T) of the rs4370946 rSNP generates three conserved punitive TFBS between the two alleles while the common C allele generates an additional 12 unique TFBS and the minor T allele generates an additional five unique TFBS. Of the 12-punitive unique TFBS generated by the C allele, the nuclear factor of kappa light polypeptide gene enhancer in B-cells 1 (NFKB1) BS is also found in the HaploReg database for this rSNP (Table). NFKB1 is a pleiotropic transcription factor present in almost all cell types and is the endpoint of 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.

The v-akt murine thymoma viral oncogene homolog 3 (*AKT3*) gene which is one of three isoforms of the AKTs which are major downstream targets of growth factor receptor tyrosine kinases that signal through PI3K, has four (rs4590656, rs12031994, rs10157763 and rs2125230) rSNPs whose punitive TFBS [9] were analyzed for binding site verification with the HaploReg and RegulomeDB databases (Table). The rs4590656 and rs12031994 rSNPs generate two and six conserved punitive TFBS, respectively (Table). The forkhead box A 1 & 2 genes were found to be conserved between both alleles (C/T) of the rs4590656 rSNP and also found in the HaploReg database for this rSNP (Table). The interferon regulatory factor 1 (IRF1) and the sex determining region Y (Sox) were found to be conserved between both alleles (G/A) of the rs12031994 rSNP and also found in the HaploReg database for this rSNP (Table). The rs10157763 rSNP generates four conserved punitive TFBS between the two (C/T) alleles and an additional five unique punitive TFBS for the common C allele and four unique TFBS for the minor T allele (Table). Of the four-unique punitive TFBS generated by the minor T allele, the CCCTC-binding factor (zinc finger protein) (CTCF) BS is also found in the HaploReg and RegulomeDB databases for this rSNP (Table). The rs2125230 rSNP generates two conserved punitive TFBS between the two (G/A) alleles and an additional four unique punitive TFBS for the G allele and nine additional unique punitive TFBS for the A allele (Table). Of the nine-unique punitive TFBS generated by the minor A allele, the forkhead box A1 (Foxa) BS is also found in the HaploReg database for this rSNP (Table).

The activating transcriptional factor 3 (*ATF3*) gene which is a member of the activating transcription factor/cAMP responsive element binding (CREB) protein family of transcription factors has one rSNP (rs11119982) whose punitive TFBS [10] was analyzed for binding site verification with the HaploReg and RegulomeDB databases (Table). The rs11119982 rSNP generates six conserved punitive TFBS and one additional unique TFBS for the common C allele and five additional unique TFBS for the minor T allele (Table). Of the five-additional unique punitive TFBS for the A allele, the natural killer 2 (NKX2) BS is also found in the HaploReg database for this rSNP (Table). The NKX2 TF acts as a negative regulator of chondrocyte maturation.

The type 2 deiodinase (*DIO2*) gene encodes a deiodinase that converts that converts the thyroid prohormone, thyroxine (T₄) to the biologically

active triiodothyronine (T₃). T₃ is involved in regulating energy balance and glucose metabolism. Six rSNPs of the *DIO2* gene whose punitive TFBS [12] were analyzed for binding site verification with the HaploReg and RegulomeDB databases (Table). The rs12885300, rs225010 and rs225011 rSNPs were found to have the TFBS PRDM1, Pax, RORa1, respectively, that were found with both alleles of each SNP. The rs225013 rSNP generates one conserved punitive TFBS and two addition unique TFBS for the common G allele and nine additional unique punitive TFBS for the minor T allele. Of the nine additional unique punitive TFBS for the T allele, the Hepatocyte Nuclear Factor 1-Alpha (HNF1a) and Paired box gene 2 (Pax) were also found in the HaploReg database for this rSNP (Table) where HNF1a is a TF that regulates the expression of several hepatic genes and Pax is a TF that plays a role in kidney cell differentiation. The rs225014 rSNP was found to have the RXRa TFBS for both alleles. The HaploReg and RegulomeDB databases indicate the CTCF, RAD21, SMC3, MAX, YY1 and cMYC TFs bind the DNA at the location of this SNP (Table). The rs6574549 rSNP was found to have seven conserved punitive TFBS and five additional unique TFBS for the common T allele and four additional unique TFBS for the minor G allele (Table1). The NK2 Homeobox 1 (Nkx2) also known as Thyroid Nuclear Factor 1 and NK3 Homeobox 1 (Nkx3) have TFBS for both alleles of this SNP. Of the five-additional unique TFBS for the common T allele of this SNP, the AT rich interactive domain 3A (Arid3a) BS was found in the HaploReg and RegulomeDB databases (Table). Of the four-additional unique TFBS for the minor G allele of the SNP, the POU class 2 homeobox 2 (Pou2f2) BS the HaploReg and RegulomeDB databases (Table).

The endothelial Per-Arnt-Sim (PAS) domain protein 1 (*EPAS1*) gene that encodes hypoxia-inducible-factor-2 alpha (*HIF1a*) is a TF involved in the response to hypoxia. Two *EPAS1* rSNPs (rs6756667 and rs7589621) whose punitive TFBS [13] were analyzed for binding site verification with the HaploReg and RegulomeDB databases (Table). The rs6756667 rSNP (A/G) was found to have 22 conserved punitive TFBS among both alleles where the Activating Transcription Factor 4 (ATF4) TFBS occurs with both alleles. The rs7589621 rSNP (G/A) was found to have 27 conserved punitive TFBS among both alleles where the Glucocorticoid modulatory element binding protein 2 (GMEB2), NK2 Homeobox 1 (Nkx2) also known as Thyroid Nuclear Factor 1 and NK3 Homeobox 1 (Nkx3) and POU class 2 homeobox 2 (Pou2f2) TFBS are found with both alleles of the SNP.

The lysosomal acid lipase A (*LIPA*) gene transcribes the lysosomal acid lipase (LAL) which hydrolyzes cholesteryl esters and triglycerides in the cell lysosome generating free cholesterol and fatty acids. Two *LIPArSNPs* (rs2246833 and rs1412444) whose punitive TFBS [14] were analyzed for binding site verification with the HaploReg and RegulomeDB databases (Table). The rs2246833 rSNP (C/T) was found to have five conserved punitive TFBS among the two alleles and three additional unique TFBS with the common C allele and one additional TFBS with the minor T allele (Table). Of the three-additional unique TFBS occurring with the common C allele, the zinc finger protein 143 (ZNF143) TFBS was also found in the HaploReg and RegulomeDB databases (Table). The CTCF TF has been found to bind this SNP location in the HaploReg and RegulomeDB databases (Table). The rs1412444 rSNP (C/T) was found to have 13 conserved punitive TFBS among the two alleles and five additional unique TFBS occurring with the common C allele and 12 additional unique TFBS occurring with the minor T allele. The ELK1, ELK4, FEV, FLI1, GABPA TFBS are found with both alleles while the protein C-ets-1 (Ets1) TFBS occurs with only the major C allele.

Signal transducer and activator of transcription 4 (*STAT4*) gene is important for signaling by interleukins (IL-12 and IL-23) and type 1 interferons. Three *STAT4* rSNPs (rs11889341, rs8179673 and rs7572482) whose punitive TFBS [15] were analyzed for binding site verification with the HaploReg and RegulomeDB databases (Table). The rs11889341 rSNP (C/T) was found to have seven conserved punitive TFBS among the two alleles and three additional unique TFBS with the common C allele and 15 unique TFBS with the minor T allele (Table). Of the 15-additional unique TFBS occurring with the common T allele, the Foxo1 TFBS was also found in the HaploReg and RegulomeDB databases (Table). The Foxo1 TF is the main target of insulin signaling and regulates metabolic homeostasis in response to oxidative stress. The rs8179673 rSNP (T/C) was found to have eight conserved punitive TFBS among the two alleles and nine additional unique TFBS with the common T allele and 15 unique TFBS with the minor C allele (Table). Among the conserved TFBS for both alleles, the Forkhead box (Fox), SRY (sex determining region Y) (Sox) and TATA Box binding protein (TBP) were also found in the HaploReg and RegulomeDB databases (Table). Of the nine unique TFBS for the common T allele, the AT rich interactive domain 3A (Arid3a) TFBS was also found in HaploReg and RegulomeDB databases (Table). Of the 15 unique TFBS for the minor C allele, the hepatocyte nuclear factors 1 & 4 (HNF1 & 4) were also found in the HaploReg and RegulomeDB databases (Table). HNF1 & 4 regulates the tissue specific expression of multiple genes, especially in pancreatic islet cells and liver. The rs7572482 rSNP (A/G) was found to have four conserved punitive TFBS among the two alleles and eight unique TFBS with the common A allele and four unique TFBS with the minor G allele. Of the unique TFBS for the A allele, the POU Class 5 Homeobox 1 (POU5F1) and SRY (sex determining region Y) (Sox) TFs form a trimeric complex on DNA and controls the expression of a number genes involved in embryonic development. These TFBS were also found in the HaploReg and RegulomeDB databases (Table). Also among the unique TFBS for the A allele, the early B-cell factor 1 (EBF1) has a BS for the bound protein as reported in the HaploReg and RegulomeDB databases (Table).

The thromboxane A2 receptor (*TBXA2R*) gene regulates different downstream signaling cascades and induces many cellular responses including the intracellular calcium influx, cell migration, proliferation and apoptosis. Three *TBXA2R* rSNPs (rs2238633, rs2238634 and rs4523) whose punitive TFBS [26] were analyzed for binding site verification with the HaploReg and RegulomeDB databases (Table). The rs2238633 rSNP (G/T) was found to have two conserved punitive TFBS among the two alleles and two unique TFBS for the common G allele and two unique TFBS with the minor T allele (Table). The Krueppel-like factor 4 (Klf4) TFBS was found to occur with both alleles. The Klf4 TF regulates the expression of key TFs during embryonic development. The rs2238634 rSNP (G/T) was found to have one conserved punitive TFBS among the two alleles and five unique TFBS with the minor T allele (Table). The zinc finger X-chromosomal protein (ZFX) TFBS was found to occur with both alleles. The ZFX TF is as member of Krueppel C2H2-type zinc-finger protein family and acts as a probable transcription activator. Of the five unique TFBS found with the minor T allele, the hepatocyte nuclear factor 4 (HNF4) BS was found in the HaploReg database (Table) as well as evidence for the HNF4 bound protein to this rSNP site. HNF4 regulates the tissue specific expression of several hepatic genes. The rs4523 rSNP (T/C) was found to have five conserved punitive TFBS among the two alleles and four unique TFBS for the common T allele and five unique TFBS for the minor C allele. The protein C-ets-1 (Ets1) TFBS was found to occur among both alleles. The Ets1 TF has been shown to interact with

Tyrosyl-DNA Phosphodiesterase 2 (TTRAP), Ubiquitin Conjugating Enzyme E2 I (UBE2I) and death associated proteins. Of the four unique TFBS for the common T allele, the nuclear receptor subfamily 3 group C member 1 (GR) TFBS was found in the HaploReg database (Table). The protein acts as a regulator of other TFs and affects inflammatory responses, cellular proliferation and differentiation in target tissues.

The vascular endothelial growth factor A (VEGFA) gene transcribes a signaling protein involved in the regulation of angiogenesis, vasculogenesis and endothelial cell growth. Six VEGFA rSNPs (rs699947, rs79469752, rs13207351, rs28357093, rs1570360 and rs2010963) whose punitive TFBS [18, 20, 27, 28] were analyzed with the HaploReg and RegulomeDB databases (Table). The rs699947 rSNP (C/A) was found to create two conserved punitive TFBS for both alleles and three unique TFBS with common C allele and one unique TFBS with the minor A allele (Table). The transcription factor CP2-like 1 (Tcfcp2l1) BS occurs with both of the alleles (Table) and the TF acts as a suppressor. The rs79469742 rSNP (C/T) was found to create one conserved punitive TFBS for both alleles and three unique TFBS with the common C allele and six unique TFBS with the minor T allele (Table). The paired box 5 (Pax5) TFBS occurs with both of the alleles (Table) and the TF is an important regulator in early development, and alterations in the expression of the gene are thought to contribute to neoplastic transformation. The TF is found to be bound at this rSNP site as indicated in the HaploReg and RegulomeDB databases (Table). The rs13207351 rSNP (G/A) was found to create one conserved punitive TFBS for both alleles and one additional unique TFBS for the common G allele and two additional unique TFBS for the minor A allele (Table). The nuclear respiratory factor 1 (Nrf1) TFBS occurs with both of the alleles (Table). The TF activates the expression of key metabolic genes regulating cellular growth and nuclear genes required for respiration, heme biosynthesis, mitochondrial DNA transcription and replication. Of the two-unique punitive TFBS occurring with the minor A allele, the paired box 5 (Pax5) TFBS was also found in the HaploReg and RegulomeDB databases (Table). The rs283570360 rSNP (A/C) was found to create one conserved punitive TFBS for both alleles and two additional unique TFBS for the common A allele and two additional unique TFBS for the minor C allele (Table). The nuclear respiratory factor 1 (Nrf1) TFBS was found to occur with both of the alleles (Table). Of the two-additional unique TFBS occurring with the common A allele, the E2F transcription factor 6 (E2F6) TFBS was found in the HaploReg and RegulomeDB databases (Table). The E2F6 TF has been found bound to the DNA at this rSNP site as reported in the HaploReg and RegulomeDB databases (Table). The TF plays a 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 rs1570360 rSNP (G/A) was found to create five conserved punitive TFBS for both alleles and three additional unique TFBS for the common G allele and six additional unique TFBS for the minor A allele (Table). The specificity protein 1 (SP1) TF has been reported to be bound to its TFBS for both alleles at this rSNP site as reported in the HaploReg and RegulomeDB databases (Table). The SP1 TF can activate or repress transcription in response to physiological and pathological stimuli. It regulates the expression of a large number of genes involved in a variety of processes such cell growth, apoptosis, differentiation and immune responses. Of the three-additional unique TFBS of the common G allele, the early growth response 1 (EGR1) TF binds the DNA at the G-allele rSNP site as reported in the HaploReg and RegulomeDB databases (Table). The TF plays a role in differentiation and mitogenesis. The rs2010963 rSNP (C/G) was found to create one conserved punitive TFBS for both alleles and three additional unique

TFBS for the common C allele and one additional unique TFBS for the minor G allele (Table). Of the three-additional unique TFBS for the common C allele, the interferon regulatory factor 1 (IRF1) TFBS was also found in the HaploReg and RegulomeDB databases (Table). The TF has been shown to play roles in regulating apoptosis and tumor-suppression.

Discussion

Non-coding SNPs identified by GWAS that have shown to be significantly associated with human disease or sickness are considered risk-associated SNPs[4]. Those near a gene causing changes in gene expression levels are considered regulatory (r) SNPs [29]. Such SNPs located in TFBS can alter the binding ability of the respective TF. There have been many reports on the possible outcome of such alterations by identifying punitive TFBS based on the two alleles of the rSNP associated with a disease or sickness [9-20]. However, such studies can only predict the possibilities of punitive TFBS associated with the disease. The HaploReg[21]and RegulomeDB[22]databases were screened for supporting evidence of which TF motif (s) reported in the previous studies that may be responsible for alteration in gene expression. The results of this investigation are provided in the Table.

As an example, the *ADRBK1* gene which has been associated with cardiovascular disease in the black population has a 3'UTR rSNP rs4370946 whose common C allele generates a NFkB1 (nuclear factor of kappa light polypeptide gene enhancer in B-cells 1) TFBS but not with the minor T allele [11]is also reported as having a motif change in HaploReg(Table). The NFkB1 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. The elimination of the NFkB1 binding site with the *ADRBK1* rs4370946-T allele could very well attribute to cardiovascular disease in the black population.

Another example, is the *AKT3* gene which has been associated with aggressive prostate cancer has an intron one rSNP rs10157761 whose minor T allele generates a CTCF (CCCTC-binding factor) TFBS but does not occur with the common C allele [9] is also reported as having a motif change in both the HaploReg and RegulomeDB databases (Table). The CTCF is a transcriptional regulatory protein with 11 highly conserved zinc finger (ZF) domains that is able to use different combinations of the ZF domains to bind DNA target sequences and proteins. CTCF serves as an insulator that interferes with the interaction between an enhancer and a promoter [30]. The elimination of the CFCT binding site with the *AKT3* rs10157761-C allele could be associated with prostate cancer. Other examples can be obtained from the Table.

In conclusion, the matching of experimental evidence from previous research as found in the HaploReg and RegulomeDB databases with computational studies that derive the names of TFBS created by rSNP alleles [9-20] leads to the most likely TFBS and causative rSNP allele responsible for disease or sickness. This process would narrow the focus of research to a small number or a single TF responsible for gene expression alternation leading to disease or sickness.

References

- Oetting WS, Bérout C, Brenner SE, Greenblatt M, Karchin R, et al. (2017) Non-Coding Variation: The 2016 Annual Scientific Meeting of the Human Genome Variation Society. *Hum Mutat*.[\[crossref\]](#)

- Hindorf 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.[\[crossref\]](#)
- 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 genetics*9: p. e1003201. [\[crossref\]](#)
- Tak YG, PJ Farnham (2015) Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. *Epigenetics Chromatin*8: p. 57. [\[crossref\]](#)
- Knight JC (2003) Functional implications of genetic variation in non-coding DNA for disease susceptibility and gene regulation. *Clin Sci (Lond)* 104: 493-501.[\[crossref\]](#)
- Wang X, Tomso DJ, Liu X, Bell DA (2005) Single nucleotide polymorphism in transcriptional regulatory regions and expression of environmentally responsive genes. *ToxicolAppPharmacol*207(2 Suppl): p. 84-90. [\[crossref\]](#)
- 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. [\[crossref\]](#)
- Buroker NE(2016) Identifying changes in punitive transcriptional factor binding sites from regulatory single nucleotide polymorphisms that are significantly associated with disease or sickness. *World J Hematology*5: 75-87.
- Buroker NE (2013) AKT3 rSNPs, Transcriptional Factor Binding Sites and Human Disease. *Open Journal of Blood Diseases*3: 116-129.
- Buroker NE (2013) ATF3 rSNPs, transcriptional factor binding sites and human etiology. *Open Journal of Genetics*3: 253-261.
- Buroker NE (2014) ADRBK1 (GRK2) rSNPs, Transcriptional Factor Binding Sites and Cardiovascular Disease in the Black Population. *Journal of Cardiovascular Disease* 2.
- Buroker NE (2014) DIO2 rSNPs, transcription factor binding sites and disease. *British Journal of Medicine & Medical Research* 9: 1-24.
- Buroker NE (2016) Computational EPAS1 rSNP analysis, transcriptional factor binding sites and high altitude sickness or adaptation. *Journal of Proteomics and Genomics research*1: 31-59.
- Buroker NE (2015) LIPA rSNPs (rs1412444 and rs2246833), Transcriptional Factor Binding Sites and Disease. *British Biomedical Bulletin* 3: 281-294.
- Buroker NE (2016) Computational STAT4 rSNP analysis, transcriptional factor binding sites and disease. *Bioinformatics and Diabetes*1: 1-36.
- Buroker NE (2014) VEGFA rSNPs, transcriptional factor binding sites and human disease. *J Physiol Sci*64: 73-76.[\[crossref\]](#)
- Buroker NE (2015) VEGFA SNPs (rs34357231 & rs35569394), Transcriptional Factor Binding Sites and Human Disease. *British Journal of Medicine & Medical Research*10: 1-11.
- BurokerNE, Ning X-H, Li K, Zhou Z-N, Cen W-J, et al. (2015) SNPs, Linkage Disequilibrium and Transcriptional Factor Binding Sites Associated with Acute Mountain Sickness among Han Chinese at the Qinghai-Tibetan Plateau. *International Journal of Genomic Medicine* 3.
- 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. *Journal of Physiological Sciences*, 63: 183-193. [\[crossref\]](#)
- Buroker NE (2014) SNP (rs1570360) in Transcriptional Factor Binding Sites of the VEGFA Promoter is Associated with Hypertensive Nephropathy and Diabetic Retinopathy. *Austin Journal of Endocrinology and Diabetes*2: 5.
- Ward LD, M Kellis(2012) HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* 40: 930-4. [\[crossref\]](#)
- Boyle AP1, Hong EL, Hariharan M, Cheng Y, Schaub MA, et al. (2012) Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 22: 1790-1797.[\[crossref\]](#)
- 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*36:102-6. [\[crossref\]](#)
- 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: D91-4. [\[crossref\]](#)

25. Sandelin A, WW Wasserman, B Lenhard (2004) ConSite: web-based prediction of regulatory elements using cross-species comparison. *Nucleic Acids Res* 32.
26. Buroker NE (2014) TBXA2R rSNPs, Transcriptional Factor Binding Sites and Asthma in Asians. *Open Journal of Pediatrics* 4.
27. Buroker NE (2014) VEGFA rSNPs, transcriptional factor binding sites and human disease. *J Physiol Sci* 64: 73-76. [[crossref](#)]
28. Buroker NE (2014) ADRBD1 (GRK2), TBXA2R and VEGFA rSNPs in KLF4 and SP1 TFBS Exhibit Linkage Disequilibrium. *Open Journal of Genetics* 4.
29. Bahreini A, Levine K, Santana-Santos L, Benos PV, Wang P et al. (2016) Non-coding single nucleotide variants affecting estrogen receptor binding and activity. *Genome Med* 8:1.
30. Ong CT, Corces VG (2014) CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet* 15: 234-246. [[crossref](#)]

Copyright: ©2017 Buroker NE. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.