OPEN ACCESS
Freely Available Online

JOURNAL OF BIOINFORMATICS AND DIABETES
ISSN NO: 2374-9431

Research Article

DOI : 10.14302/issn.2374-9431.jbd-15-890

Computational STAT4 rSNP analysis, transcriptional factor binding sites and disease

Norman E. Buroker[1]

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

Abstract

Purpose - Signal Transducer and Activator of Transcription 4 (*STAT4)* is important for signaling by interleukins (IL-12 and IL-23) and type 1 interferons and has been found to have several simple nucleotide polymorphisms (SNPs) associated with human disease. *STAT4* SNPs were computationally examined with respect to changes in potential transcriptional factor binding sites (TFBS) and these changes were discussed in relation to human disease.

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 *STAT4* gene from 4 kb upstream of the transcriptional start site to 8.3 kb past the 3'UTR. The JASPAR CORE database was also involved in computing each nucleotide occurrence (%) within the TFBS.

Results - The *STAT4* SNPs in the 70 kb intron between exon 2 and 3 are in linkage disequilibrium and have previously been found to be significantly associated with several vasculitis diseases as well as diabetes. 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 *STAT4* gene. These *STAT4* 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 *STAT4* regulation. These regulatory changes were discussed with respect to changes in human health that result in disease.

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

E-mail: nburoker@u.washington.edu

Keywords: *STAT4,* rSNPs, TFBS, disease

Received : Dec 15, 2015 **Accepted :** Feb 11, 2016 **Published :** Feb 26, 2016

INTRODUCTION

The Janus Kinase-Signal Transducers and Activators of Transcription (JAK-STAT) pathways play a critical role in immune, neuronal, hematopoietic and hepatic systems [1]. JAK-STAT is a principal signal transduction pathway in cytokine and growth factor signaling as well as regulating various cellular processes such as cell proliferation, differentiation migration and survival [2]. JAK-STAT provides the principle intracellular signaling mechanism required for a wide array of cytokines [3, 4]. The STAT portion of the signaling cascade has seven mammalian family members which are STAT1, 2, 3, 4, 5a, 5b and 6 [3, 4]. These STATs bind thousands of transcriptional factor binding sites (TFBS) in the genome and regulate the transcription of many protein-coding genes, miRNAs and long noncoding RNAs [4]. The *STAT 4* gene which is important for signaling by interleukins (IL-12 and IL-23) and type 1 interferons [4] has been found to have several simple nucleotide polymorphisms (SNPs) associated with human disease [5-12]. STAT4 transduces IL-12, IL-23 and type 1 interferon-mediated signals into helper T (Th) cells (Th1 and Th17) differentiation, monocyte activation, and interferon-gamma production [12, 13]. The STAT4-dependent cytokine regulation is found in the pathogenesis of autoimmune disease [14, 15] such as systemic lupus erythematosus (SLE), rheumatoid arthritis (RA) and inflammatory bowel disease (IBD) [16, 17].

The *STAT4* gene maps to human chromosome 2q32.3 and is about 143 kb in size. The coding region consists of 22 exons with a large 70 kb intron between exons 2 and 3. Several SNPs in the gene have been significantly association with Behcet's Disease [18], diabetes risk [11], hepatitis B virus-related hepatocellular carcinoma [6, 10, 19, 20], inflammatory bowel disease [21], juvenile idiopathic arthritis [22], primary biliary cirrhosis and Crohn's disease [23], severe renal insufficiency in lupus nephritis [8], systemic lupus erythematosus [5] and ulcerative colitis [24] (Table 1). The rs7574865 *STAT4* SNP has been found to be significantly associated with diabetes [11], hepatitis B virus-related hepatocellular carcinoma [6, 10, 19, 20], inflammatory bowel disease [21], juvenile idiopathic arthritis [22], primary biliary cirrhosis and Crohn's disease [23], severe renal insufficiency in lupus nephritis [8], systemic lupus erythematosus [5] and ulcerative colitis [24]. The rs11889341 *STAT4* SNP has been found to be significantly associated with diabetes [11], hepatitis B virus (HBV) infection, HBV-related cirrgisus and hepatocellular carcinoma [23], severe renal insufficiency in lupus nephritis [8], and systemic lupus erythematosus [5]. The rs8179673 *STAT4* SNP has been found to be significantly associated with diabetes [11], hepatitis B virus (HBV) infection, HBV-related cirrgisus and hepatocellular carcinoma [23] and systemic lupus erythematosus [5]. The rs7582694 *STAT4* SNP has been found to be significantly associated with hepatitis B virus (HBV) infection, HBV-related cirrgisus and hepatocellular carcinoma [23] and severe renal insufficiency in lupus nephritis [8]. The rs7574070 and rs7572482 *STAT4* SNPs have been found to be significantly associated with Behcet's disease [18]. The rs7572482 *STAT4* SNP is located in the promoter region while the remaining SNPs are located in the large 70 kb intron between exon 2 and 3. The reports listed above indicate that these SNPs are in strong linkage disequilibrium (LD) with each other.

Single nucleotide changes that affect gene expression by impacting gene regulatory sequences such as promoters, enhances, and silencers are known as regulatory SNPs (rSNPs) [25-28]. A rSNPs within a transcriptional factor binding site (TFBS) can change a transcriptional factor's (TF) ability to bind its TFBS [29-32] in which case the TF would be unable to effectively regulate its target gene [33-37]. This concept is examined for the above *STAT4* rSNPs and their allelic

(Continued on page 21)

Table 1. *STAT4* SNPs and disease. The SNPs have been found to be significantly associated with these diseases.

The SNPs are located in *STAT4* intron 3. MAF is the minor allele frequency. LD is linkage disequilibrium.

Disease	SNP	Chr 2 Pos	Alleles	MAF	Risk Allele	LD	Study Group	Reference
Behcet's	rs7574070 rs7572482	191145762 191150346	A/C A/G	C=0.47 A=0.47		Yes Yes	Chinese	7
Diabetes	rs11889341 rs7574865 rs8179673 rs10181656	191079016 191099907 191104615 191105153	C/T G/T T/C C/G	T=0.34 T=0.25 C=0.26 G=0.26		Yes Yes Yes Yes	Asian, Caucasian	11
Hepatitis B virus-related hepatocellular carcinoma	rs7574865	191099907	G/T	T=0.25	G		Chinese	6
HBV infection, HBV-related cirrgisus and hepatocellular carcinoma	rs7574865 rs7582694 rs11889341 rs8179673	191099907 191105394 191079016 191104615	G/T G/C C/T T/C	T=0.25 C=0.33 T=0.34 C=0.26	G G C T	Yes Yes Yes Yes	Chinese	10
HBV viral clearance	rs7574865	191099907	G/T	T=0.25	G		Tibetan, Uygur	19
Hepatocellular carcinoma	rs7574865	191099907	G/T	T=0.25	G		Korean	20
Inflammatiory bowel disease	rs7574865	191099907	G/T	T=0.25	G		Chinese, Causcians	21
Juvenil Idiopathic arthritis	rs7574865	191099907	G/T	T=0.25	T		Han Chinese	22
Primary biliary cirrhosis and Crohn's disease	rs7574865	191099907	G/T	T=0.25	T		Japanese	23
Severe renal insufficiency in lupus nephritis	rs11889341 rs7574865 rs7568275 rs7582694	191079016 191099907 191101726 191105394	C/T G/T C/G G/C	T=0.28 T=0.23 G=0.28 C=0.22		Yes Yes Yes Yes	Swedish	8
Systemic Lupus Erythematosus	rs11889341 rs7574865 rs8179673 rs10181656	191079016 191099907 191104615 191105153	C/T G/T T/C C/G	T=0.28 T=0.23 C=0.26 G=0.26		Yes Yes Yes Yes	European descent	5
Ulcerative colitis	rs7574865	191099907	G/T	T=0.25	T		European descent	24

association with TFBS, where computation analyses [38-41] was used to identify TFBS alterations created by the *STAT4* rSNPs. Recent reports have also introduced the concept of modeling of epigenetic modifications to transcriptional factor binding sites in the control of gene expression [42, 43]. In this report, the rSNP associations with changes in potential TFBS are discussed with their possible relationship to these diseases in humans.

METHODS

The JASPAR CORE database [44, 45] and ConSite [46] 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 [47]. The Vector NTI Advance 11.5 computer program (Invitrogen, Life Technologies) was used to locate the TFBS in the *STAT4* gene (NCBI Ref Seq NM_003151) from 4 kb upstream of the transcriptional start site to 8.3 kb past the 3'UTR which represents a total of 130.9 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 2 & Appendix).

RESULTS

STAT4 rSNPs and TFBS

The *STAT4* gene transcribes the transcriptional factor (TF) protein which is part of a family of STAT TFs that act as transcriptional activators in response to cytokines and growth factors. This protein is essential for mediating responses to IL12 in lymphocytes, and regulating the differentiation of T helper cells. Due to the importance of this gene in signal transduction and activation of transcription, *STAT4* SNPs associated with disease were computationally evaluated with regard to TFBS. The rs7574865 *STAT4* SNP located in the large 70 kb intron has been found to have the most significant association with human disease (Table 1).

The common rs7574865 SNP *STAT4*-G allele creates three unique TFBS for the FOXL1, MAX and ZNF354C TFs, which are involved with the regulation of metabolism, cell proliferation and gene expression during ontogenesis, transcription regulation and repression, respectively (Table 2, Appendix). The minor *STAT4*-T allele creates eight unique TFBS for the ARID3A, FOXQ1, NKX2-5, NR1HE::RXR1, PDX1, PRRX2, SOX5 and SRY TFs which are involved with the control of cell cycle progression, differentiation of lung

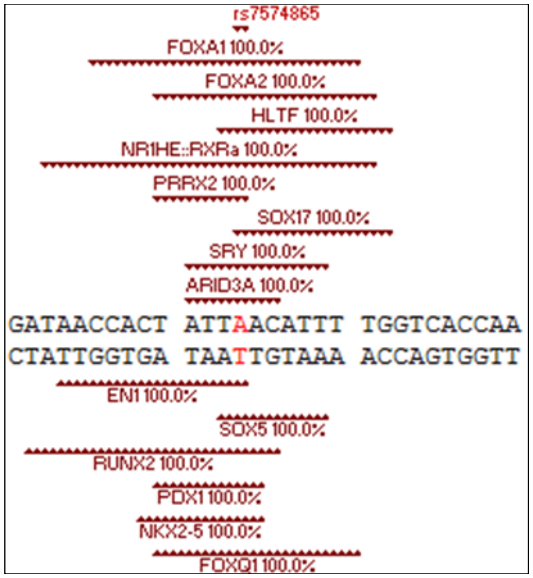

Figure 1. Double stranded DNA from the *STAT4* gene showing the potential TFBS for fourteen different TFs which can bind their respective DNA sequence either above (+) or below (-) the duplex (cf. Table 2). The rs7574865 rSNP minor *STAT4*-T allele is found in each of these TFBS. As shown, this rSNP is located in the 70 kb intron between exon 2 and 3 of the *STAT4* gene. Also included with the potential TFBS is their % sequence homology to the duplex.

Table 2. The *STAT4* 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.

SNP	Allele	TFs	Protein name	# of Sites	TFBS	Strand
rs7574865	G	EN1	Engrailed homeobox 1	1	gaatagtggtt	plus
					g=20%	
		FOXA1	Forkhead box A1	1	ccacTaTTcaCattt	minus
					c=0%	
		FOXA2	Forkhead box A2	1	TaTTcACatttt	minus
					c=0.1%	
		FOXL1	Forkhead box L1	8	tgtgaATA	plus
					g=17%	
		HLTF	Helicase-like transcription factor	1	tcaCaTtttg	minus
					c=20%	
		MAX	MYC Associated Factor X	3	attCACaTtt	minus
					C=100%	
		RUNX2	Runt-related transcription factor 2	1	gtgaataGTGGttat	plus
					g=40%	
		SOX17	SRY (sex determining region Y)-box 17	2	cacATTtTg	minus
					c=29%	
		ZNF354C	Zinc finger protein 354C	81	attCAC	minus
					C=100%	
	T	ARID3A	AT rich interactive domain 3A	81	ATtAAc	minus
	0.25		(BRIGHT-like)		A=100%	
		EN1	Engrailed homeobox 1	1	taatagtggtt	plus
					t=30%	
		FOXA1	Forkhead box A1	1	ccacTaTTaaCattt	minus
					a=0%	

Table 2 Continued

SNP	Allele	TFs	Protein name	# of Sites	TFBS	Strand
		HLTF	Helicase-like transcription factor	1	taaCaTtttg	minus
					a=34%	
		NKX2-5	Natural killer 3 homeobox 2	23	ttAAtag	plus
					t=76%	
		NR1HE: RXRa	Nuclear Receptor Subfamily 1, Group H, Member 3	1	TaaccactatTaacatttt	minus
			Retinoid X receptor, alpha		a=44%	
		PDX1	Pancreatic and duodenal homeobox 1	158	tTAATa	plus
					T=100%	
		PRRX2	Paired related homeobox 2	518	tATTA	minus
					A=98%	
		RUNX2	Runt-related transcription factor 2	1	gttaataGTGGttat	plus
					t=38%	
		SOX5	SRY (sex determining region Y)-box 5	42	AaTGTTa	plus
					T=91%	
		SOX17	SRY (sex determining region Y)-box 17	4	aacATTtTg	minus
					a=23%	
		SRY	Sex determining region Y	4	attaACAtt	minus
					a=64%	
rs11889341	C	AR	Androgen Receptor	1	aaGaAtAagatGttc	minus
					G=100%	
		CEBPb	CCAAT/enhancer binding protein (C/EBP),	1	tcTTttaccAc	plus
					c=6%	
		FOXL1	Forkhead box L1	24	aaagaATA	minus
					g=26%	
		HLTF	Helicase-like transcription factor	2	attCtTttac	plus
					C=100%	
		NR3C1	Nuclear Receptor Subfamily 3, Group C,	1	agaataagatGTtCa	minus
			(Glucocorticoid Receptor)		g=80%	
		SOX3	SRY (sex determining region Y)-box 3	1	ttaTTcTttt	plus
					c=7%	

Table 2 Continued

SNP	Allele	TFs	Protein name	# of Sites	TFBS	Strand
		SOX5	SRY (sex determining region Y)-box 5	73	aTTcTTt	plus
					c=0%	
		SOX6	SRY (sex determining region Y)-box 56	6	ttaTTcTttt	plus
					c=0%	
		SRY	Sex determining region Y	9	taaaAgAAa	minus
					g=0%	
		ZNF263	Zinc finger protein 263	1	agaGcAgtggtaaaagaataa	minus
					g=75%	
	T	CDX2	Caudal type homeobox 2	13	tggtaAaaAAa	minus
	0.34				A=100%	
		FOXA1	Forkhead box A1	1	atctTaTTtttttac	plus
					t=0%	
		FOXD1	Forkhead Box D1	14	gTAAAaAa	minus
					A=100%	
		FOXD3	Forkhead box D3	29	tctTaTTtttt	plus
					t=68%	
		FOXI1	Forkhead box I1	29	tctTaTTTtttt	plus
					T=100%	
		FOXL1	Forkhead box L1	59	aaaaaATA	minus
					a=57%	
		FOXO1	Forkhead Box O1	13	attTtTTtacc	plus
					T=100%	
		FOXO1	Forkhead Box O1	2	tctTaTTtttt	plus
					t=88%	
		FOXO3	Forkhead Box O3	9	ggtAAAaA	plus
					A=92%	
		FOXP1	Forkhead box P1	1	cagtggtAAAaAaat	minus
					A=100%	

Table 2 Continued

SNP	Allele	TFs	Protein name	# of Sites	TFBS	Strand
		FOXP2	Forkhead box P2	13	tggTAAAaAaa	minus
					A=100%	
		GATA1	GATA binding protein 1	2	atcTTATtttt	plus
					t=88%	
		GATA1	GATA binding protein 1	10	tttTTAccact	plus
					t=51%	
		GATA2	GATA binding protein 2	2	atttttTTAcCact	plus
					t=54%	
		GATA2	GATA binding protein 2	1	aacatcTTATtttt	plus
					t=72%	
		GATA3	GATA Binding Protein 3	24	AaATAAga	minus
					A=100%	
		GATA4	GATA binding protein 4	2	tcTTATttttt	plus
					t=79%	
		HLTF	Helicase-like transcription factor	1	catCtTattt	plus
					t=25%	
		MEF2C	Myocyte Enhancer Factor 2C	1	ggtaaaaAAATAaga	minus
					A=95%	
		MEF2C	Myocyte Enhancer Factor 2C	1	gtggtaaAAAaAtaa	minus
					A=97%	
		NR3C1	Nuclear Receptor Subfamily 3, Group C,	1	aaaataagatGTtCa	minus
			(Glucocorticoid Receptor)		a=15%	
		SOX5	SRY (sex determining region Y)-box 5	164	aTTtTTt	plus
					t=0%	
		SOX6	SRY (sex determining region Y)-box 56	8	ttaTTtTttt	plus
					t=0%	
		SRY	Sex determining region Y	19	taaaAaAAt	minus
					a=0%	

Table 2 Continued

SNP	Allele	TFs	Protein name	# of	TFBS	Strand
		SRY	Sex determining region Y	4	gtaaAaAAa	minus
					A=100%	
		ZNF263	Zinc finger protein 263	1	agaGcAgtggtaaaaaaataa	minus
					a=19%	
rs8179673	T	ARID3A	AT rich interactive domain 3A	397	AatAAa	plus
			(BRIGHT-like)		t=63%	
		ARID3A	AT rich interactive domain 3A	227	ATttAa	minus
			(BRIGHT-like)		A=100%	
		EN1	Engrailed homeobox 1	1	aaataaaggtc	plus
					t=80%	
		FOXA1	Forkhead box A1	1	ccTTTaTTtaatata	minus
					a=27%	
		FOXL1	Forkhead box L1	25	attaaATA	plus
					T=91%	
		FOXL1	Forkhead box L1	23	atttaATA	minus
					a=30%	
		GATA3	GATA Binding Protein 3	27	AaATAAag	plus
					T=100%	
		HOXA5	Hoxa5	27	ctttatTt	minus
					a=88%	
		LHX3	LIM homeobox 3	3	atATTAAaTaaag	minus
					T=95%	
		NFIL3	Nuclear factor, interleukin 3 regulated	2	TTAttTAAtat	minus
					A=96%	
		NKX3-1	NK3 homeobox *1*	72	tTAtTTA	minus
					A=100%	

Table 2 Continued

SNP	Allele	TFs	Protein name	# of	TFBS	Strand
		RORA_1	RAR-related orphan receptor A	11	ataaaGGTCc	plus
					t=60%	
		RXRa	Retinoid X receptor, alpha	5	taaAGgtCcat	plus
					t=5%	
		SOX2	SRY (sex determining region Y)-box 2	11	CCtTTaTt	minus
					a=0%	
		SOX3	SRY (sex determining region Y)-box 3	1	cctTTaTtta	minus
					a=0%	
		SOX6	SRY (sex determining region Y)-box 6	1	cctTTaTtta	minus
					a=0%	
		SOX10	SRY (sex determining region Y)-box 10	142	cttTaT	minus
					a=0%	
		SRY	Sex determining region Y	9	ttaaAtAAa	plus
					t=7%	
		TBP	TATA Box Binding Protein	1	gtATAtAttaaataa	plus
					t=16%	
	C	BRCA1	breast cancer 1, early onset	39	acAaagg	plus
	0.26				c=81%	
		FOXA1	Forkhead box A1	1	ccTTTgTTtaatata	minus
					g=9%	
		FOXA2	Forkhead box A2	1	TgTTtaatatat	minus
					g=72%	
		FOXA2	Forkhead box A2	1	TaTggACctttg	minus
					g=36%	
		FOXD1	Forkhead Box D1	6	tTAAACAa	plus
					C=90%	

Table 2 Continued

SNP	Allele	TFs	Protein name	# of Sites	TFBS	Strand
		FOXH1	Forkhead Box H1	1	tatAtTaaACa	plus
					C=100%	
		FOXO1	Forkhead Box O1	1	cttTGTTtaat	minus
					G=100%	
		FOXP1	Forkhead box P1	1	ata-	plus
					c=89%	
		FOXP2	Forkhead box P2	1	tatTAAACAaa	plus
					C=99%	
		FOXQ1	Forkhead box Q1	1	ctttGTTTAat	minus
					G=100%	
		HLTF	Helicase-like transcription factor	1	gacCtTtgtt	minus
					g=17%	
		HNF1A	Hepatocyte Nuclear Factor 1 homeobox A	1	gtTTAaTatatact	minus
					g=67%	
		HNF4A	Hepatocyte Nuclear Factor 4, Alpha	1	tggacctttgtttaa	minus
					g=12%	
		HNF4G	Hepatocyte Nuclear Factor 4, Gamma	1	at-	plus
					C=93%	
		JUN::FOS	Jun Proto-Oncogene	48	TtAaacA	plus
			FBJ Murine Osteosarcoma Viral Oncogene		c=83%	
		RXRa	Retinoid X receptor, alpha	1	caaAGgtCcat	plus
					c=85%	
		SPX2	SRY (sex determining region Y)-box 2	2	CCtTgGTc	minus
					G=100%	
		SOX3	SRY (sex determining region Y)-box 3	1	cctTTGTttA	minus
					G=93%	

Table 2 Continued

SNP	Allele	TFs	Protein name	# of	TFBS	Strand
		SOX5	SRY (sex determining region Y)-box 5	134	tTTGTTt	minus
					G=96%	
		SOX6	SRY (sex determining region Y)-box 56	1	cCtTTGTtta	minus
					G=100%	
		SOX9	SRY (sex determining region Y)-box 9	4	cctTtGttt	minus
					G=95%	
		SOX10	SRY (sex determining region Y)-box 10	141	cttTgT	minus
					g=86%	
		SRY	Sex determining region Y	4	ttaaACAAa	plus
					C=93%	
		TBP	TATA Box Binding Protein	1	gtATAtAttaaacaa	plus
					c=30%	
rs7582694	G	BATF::JUN	Basic leucine zipper transcription factor,	1	tctaTGtgTcA	minus
			ATF-like Jun proto-oncogene		c=5%	
		CEBPa	CCAAT/enhancer binding protein (C/EBP),	1	gTTgCatactc	minus
					c=32%	
		FOSL2	FOS-Like Antigen 2	1	ctaTGtgTCAt	minus
					c=19%	
		FOXC1	Forkhead box C1	3	atagaGTA	plus
					g=25%	
		HLTF	Helicase-like transcription factor	1	acaCaTagag	plus
					g=37%	
		HLF	Hepatic Leukemia Factor	1	ggTtgcatactc	minus
					c=28%	
		JUN(var.2)	Jun Proto-Oncogene	2	actctaTGtgTCAt	minus
					c=10%	
		JUNB	Jun B Proto-Oncogene	1	ctaTGtgTCAt	minus
					c=20%	

Table 2 Continued

SNP	Allele	TFs	Protein name	# of	TFBS	Strand
		MAX	MYC Associated Factor X	1	tgaCACaTag	plus
					g=29%	
		SOX3	SRY (sex determining region Y)-box 3	2	tctaTGTgtc	minus
					c=75%	
		SOX10	SRY (sex determining region Y)-box 10	77	ctaTgT	minus
					c=86%	
		T	T, Brachyury Homolog	1	cTAtGTGTcAt	minus
					c=70%	
	C	BATF: JUN	Basic leucine zipper transcription factor,	2	tgtaTGtgTcA	minus
	0.33		ATF-like Jun proto-oncogene		g=27%	
		BH1HE40	Basic Helix-Loop-Helix Family, Member E40	2	gaCACaTacag	plus
					c=75%	
		CEBPa	CCAAT/enhancer binding protein (C/EBP), alpha	1	gTTgCatactg	minus
					g=6%	
		FOSL2	FOS-Like Antigen 2	1	gtaTGtgTCAt	minus
					g=39%	
		FOXC1	Forkhead box C1	7	atacaGTA	plus
					c=25%	
		FOXC1	Forkhead box C1	10	atactGTA	minus
					G=100%	
		HLF	Hepatic Leukemia Factor	1	ggTtgcatactg	minus
					g=17%	
		HIF1a: ARNT	Hypoxia Inducible Factor 1, Alpha Subunit	11	gtatGTGt	minus
			Aryl Hydrocarbon Receptor Nuclear Trans-locator		g=47%	
		JUN (var.2)	Jun Proto-Oncogene	1	actgtaTGtgTCAt	minus
					g=32%	

Table 2 Continued

SNP	Allele	TFs	Protein name	# of	TFBS	Strand
		JUNB	Jun B Proto-Oncogene	1	gtaTGtgTCAt	minus
					g=24%	
rs7574070	A	CEBPB	CCAAT/enhancer binding protein (C/EBP),	1	aaTgtCtccAt	plus
					A=100%	
		MZF1_1-4	Myeloid Zinc Finger 1	112	tGGaGA	minus
					t=40%	
		NFE2L1::MafG	Nuclear Factor, Erythroid 2-Like 1	123	caTGAa	plus
			V-Maf Avian Musculoaponeurotic		a=85%	
			Fibrosarcoma Oncogene Homolog G			
		PAX2	Paired box gene 2	3	cttCatgg	minus
					t=35%	
		RFX1	Regulatory Factor X, 1	1	ttcttCatgGagAC	minus
			(Influences HLA Class II Expression)		t=84%	
		RFX5	Regulatory factor X, 5 (influences HLA	1	cttCatgGagA-	minus
			class II expression)		t=78%	
		STAT3	Signal transducer and activator of	1	cTtCatGgAga	minus
			transcription 3 (acute-phase response		t=30%	
		STAT5A::STAT5B	Signal transducer and activator of	1	gtcTCcatGAA	plus
			transcription 5A and transcription 5B		a=43%	
		THAP1	THAP domain containing,	3	tctCCatga	plus
			apoptosis associated protein 1		a=29%	
		ZNF354C	Zinc finger protein 354C	87	ctCCAt	plus
					A=100%	
	C	EBF1	Early B-cell factor 1	7	ttCttcaGgGa	minus
	0.47				G=97%	
		ELK1	ELK1, member of ETS oncogene family	2	ctccctGAag	plus
					c=86%	

Table 2 Continued

SNP	Allele	TFs	Protein name	# of Sites	TFBS	Strand
		STAT1	Signal Transducer And Activator	6	cTTCagGGAga	minus
			Of Transcription 1, 91kDa		G=96%	
		STAT3	Signal transducer and activator of	3	cTtCagGgAga	minus
			transcription 3 (acute-phase response factor)		g=45%	
		STAT4	Signal Transducer And Activator Of Transcription 4	1	cTtcaggGAgacat	minus
					g=43%	
		STAT5A::STAT5B	Signal transducer and activator of	4	gtcTCcctGAA	plus
			transcription 5A and transcription 5B		c=37%	
		TFAP2C	Transcription factor AP-2 gamma	1	ccattCttcAGgag	minus
			(activating enhancer binding protein 2 gamma)		G=99%	
		THAP1	THAP domain containing,	3	tctCCctga	plus
			apoptosis associated protein 1		c=68%	
rs7572482	A	ARID3A	AT rich interactive domain 3A	237	ATgAAa	plus
			(BRIGHT-like)		A=100%	
		EBF1	Early B-cell factor 1	2	ttCtCatGaaa	plus
					a=27%	
		EBF1	Early B-cell factor 1	1	ttttCatGaGa	minus
					t=37%	
		FOS	FBJ Murine Osteosarcoma Viral Oncogene Homolog	1	tcTttcTCAtg	plus
					A=100%	
		NFE2L1::MafG	Nuclear Factor, Erythroid 2-Like 1	79	caTGAg	minus
			V-Maf Avian Musculoaponeurotic		T=100%	
			Fibrosarcoma Oncogene Homolog G			
		NFE2L1::MafG	Nuclear Factor, Erythroid 2-Like 1	123	caTGAa	plus
			V-Maf Avian Musculoaponeurotic		a=85%	
			Fibrosarcoma Oncogene Homolog G			

Table 2 Continued

SNP	Allele	TFs	Protein name	# of	TFBS	Strand
		POU5F1::SOX2	POU Class 5 Homeobox 1	1	ctTTctcATGaaaac	plus
			SRY (Sex Determining Region Y)-Box 2		A=90%	
		RFX1	Regulatory Factor X, 1	1	tttctCatgaaaAC	plus
			(Influences HLA Class II Expression)		a=59%	
		SPIB	Spi-B transcription factor (Spi-1/PU.1 related)	52	tgaGaAA	minus
					t=37%	
		SOX3	SRY (sex determining region Y)-box 3	1	tctTTcTcat	plus
					a=1%	
		STAT3	Signal transducer and activator of	1	tTcatgagAAa	minus
			transcription 3 (acute-phase response factor)		t=60%	
		STAT4	Signal Transducer And Activator Of Transcription 4	1	tTcatgaGAAagaa	minus
					t=35%	
		STAT6	Signal Transducer And Activator Of Transcription 6	1	tctTTCtcaaGAAaa	plus
			Interleukin-4 Induced		a=18%	
		STAT6	Signal Transducer And Activator Of Transcription 6	1	gttTTCatgaGAAag	minus
			Interleukin-4 Induced		t=51%	
		STAT5A::STAT5B	Signal transducer and activator of	1	ttTctcatGAA	plus
			transcription 5A and transcription 5B		a=43%	
		STAT5A::STAT5B	Signal transducer and activator of	1	gtTTtcatGAg	minus
			transcription 5A and transcription 5B		t=5%	
	G	ARNT	Aryl Hydrocarbon Receptor Nuclear Translocator	14	cACGaG	minus
	0.47				C=100%	
		ARNT	Aryl Hydrocarbon Receptor Nuclear Translocator	14	ctCGTG	plus
					G=100%	
		ARNT::AHR	Aryl Hydrocarbon Receptor Nuclear Translocator	14	ctCGTG	plus
			Aryl Hydrocarbon Receptor		G=96%	
		SOX3	SRY (sex determining region Y)-box 3	10	tctTTcTcgt	plus
					g=1%	

Table 2 Continued

SNP	Allele	TFs	Protein name	# of Sites	TFBS	Strand
		SPI1	Spleen focus forming virus (SFFV)	1	acgagaaaGAAgtag	minus
			proviral integration oncogene spi1		c=12%	
		STAT3	Signal transducer and activator of	3	tTcacgagAAa	minus
			transcription 3 (acute-phase response		c=36%	
		STAT4	Signal Transducer And Activator Of Tran-	1	tTcacgaGAAagaa	minus
					c=52%	
		STAT6	Signal Transducer And Activator Of Tran-	3	tctTTCtcgaGAAaa	plus
			Interleukin-4 Induced		g=65%	
		STAT6	Signal Transducer And Activator Of Tran-	1	gttTTCacgaGAAag	minus
			Interleukin-4 Induced		c=37%	
		ZBTB33	Zinc Finger And BTB Domain Containing	1	ttCtCGtGaaaactg	plus
					G=100%	
rs7568275	C	FOXA1	Forkhead box A1	1	gtaaTaTTaactgaa	minus
					g=6%	
		FOXI1	Forkhead box I1	3	ataTgTTcagtt	plus
					c=0%	
		FOXL1	Forkhead box L1	10	tgaacATA	minus
					g=9%	
		HLTF	Helicase-like transcription factor	8	gaaCaTataa	minus
					g=20%	
		NKX2-5	NK2 Homeobox 5	19	ttAActg	minus
					g=65%	
		SRF	Serum Response Factor	1	actgaaCatA-	minus
			(C-Fos Serum Response Element-Binding		g=0.44	
			Transcription Factor)			
	G	BATF:	Basic leucine zipper transcription factor,	2	atgtTGAgTtA	plus
	0.28		ATF-like Jun proto-oncogene		G=99%	

Table 2 Continued

SNP	Allele	TFs	Protein name	# of	TFBS	Strand
		BATF::JUN	Basic leucine zipper transcription factor,	1	atatTaAcTcA	minus
			ATF-like Jun proto-oncogene		c=82%	
		BRCA1	Breast Cancer 1, Early Onset	43	tcAacat	minus
					c=81%	
		FOXA1	Forkhead box A1	1	gtaa-	minus
					c=33%	
		FOXA1	Forkhead box A1	1	tataTgTTgag-	plus
					g=13%	
		FOXA2	Forkhead box A2	1	TgTTgAgttaat	plus
					g=22%	
		FOXA2	Forkhead box A2	1	TaTTaActcaac	minus
					c=33%	
		Foxd3	Forkhead box D3	1	ataTgTTgagtt	plus
					g=21%	
		FOXD3	Forkhead box D3	1	taaTaTTaactc	minus
					c=26%	
		FOXH1	Forkhead Box H1	2	ttaAcTcaACa	minus
					c=70%	
		FOXI1	Forkhead box I1	1	ataTgTTgagtt	plus
					g=0%	
		FOXL1	Forkhead box L1	16	tcaacATA	minus
					c=17%	
		HLTF	Helicase-like transcription factor	1	caaCaTataa	minus
					c=17%	
		JUN::FOS	Jun Proto-Oncogene	17	TaActcA	minus
			FBJ Murine Osteosarcoma Viral Oncogene		c=83%	

Table 2 Continued

SNP	Allele	TFs	Protein name	# of	TFBS	Strand
		MAFF	V-Maf Avian Musculoaponeurotic	1	attaacTCAacAtataaa	minus
			Fibrosarcoma Oncogene Homolog F		C=94%	
		MAFK	V-Maf Avian Musculoaponeurotic	1	ttaacTCAaCAtata	minus
			Fibrosarcoma Oncogene Homolog K		C=96%	
		ZNF354C	Zinc finger protein 354C	58	ctCCAc	minus
					C=100%	
rs10181656	C	AR	Androgen Receptor	1	tgGtACAagggGtga	minus
					G=95%	
		E2F6	E2F transcription factor 6	3	ggGtGaGAaga	minus
					G=100%	
		GATA1	GATA binding protein 1	4	ctcTTcTCacc	plus
					c=22%	
		GATA2	GATA binding protein 2	1	caactcTTcTCacc	plus
					c=40%	
		GATA4	GATA binding protein 4	3	tcTTcTCaccc	plus
					c=45%	
		HLTF	Helicase-like transcription factor	7	cccCtTgtac	plus
					c=17%	
		KLF5	Kruppel-like factor 5 (intestinal)	1	ttctCaCCCc	plus
					C=100%	
		MZF1_1-4	Myeloid Zinc Finger 1	58	gGGtGA	minus
					G=90%	
		MZF1_5-13	Myeloid Zinc Finger 1	1	caAgGgtga	minus
					g=88%	
		NR1H3: RXRa	Nuclear Receptor Subfamily 1, Group H, Member 3	1	TcacccttgTaccactac	plus
			Retinoid X receptor, alpha		c=77%	

Table 2 Continued

SNP	Allele	TFs	Protein name	# of	TFBS	Strand
		SP1	Specificity Protein 1	1	ttCtcacCcct	plus
					C=80%	
		SREBF1	Sterol regulatory element binding	2	cTCAcccctt	plus
			transcription factor 1		c=88%	
		SREBF2	Sterol regulatory element binding	2	aaGgggTGAg	minus
			transcription factor 2		g=77%	
		ZNF263	Zinc finger protein 263	1	agtGg-tacaagggtgag	minus
					g=59%	
	G	BRCA1	Breast cancer 1, early onset	12	tcAgccc	plus
	0.26				g=9%	
		GATA1	GATA binding protein 1	1	ctcTTcTCagc	plus
					g=34%	
		GATA2	GATA binding protein 2	1	caactcTTcTCag	plus
					g=45%	
		GATA4	GATA binding protein 4	1	tcTTcTCagcc	plus
					g=34%	
		HLTF	Helicase-like transcription factor	1	gccCtTgtac	plus
					g=20%	
		HNF4G	Hepatocyte Nuclear Factor 4, Gamma	1	gtgg-	minus
					c=46%	
		KLF5	Kruppel-like factor 5 (intestinal)	1	tctCagCCCt	plus
					g=38%	
		MAFB	v-maf musculoaponeurotic fibrosarcoma	15	Gctgagaa	minus
			oncogene homolog B (avian)		c=80%	
		SP1	Specificity Protein 1	1	tctcagCcctt	plus
					g=43%	

Table 2 Continued

SNP	Allele	TFs	Protein name	# of	TFBS	Strand
		SREBF1	Sterol regulatory element binding	1	cTCAgccctt	plus
			transcription factor 1		g=12%	
		STAT3	Signal transducer and activator of	1	gggCtgagAAg	minus
			transcription 3 (acute-phase response		C=91%	
		THAP1	THAP domain containing,	2	cagCCcttg	plus
			apoptosis associated protein 1		g=6%	

Appendix. Transcriptional factor (TF) discriptions.

Appendix. Transcriptional factor (TF) discriptions.	
TFs	**TF discription**
AR	The protein functions as a steroid-hormone activated transcription factor. Upon binding the hormone ligand, the receptor dissociates from accessory proteins, translocates into the nucleus, dimerizes, and then stimulates transcription of androgen responsive genes. They are expressed in bone marrow, mammary gland, prostate, testicular and muscle tissues where they exist as dimers coupled to Hsp90 and HMGB proteins.
ARID3A	Transcription factor which may be involved in the control of cell cycle progression by the RB1/E2F1 pathway and in B-cell differentiation
ARNT	This gene encodes a protein containing a basic helix-loop-helix domain and two characteristic PAS domains along with a PAC domain. The encoded protein binds to ligand-bound aryl hydrocarbon receptor and aids in the movement of this complex to the nucleus, where it promotes the expression of genes involved in xenobiotic metabolism.
ARNT::AHR	The dimer alters transcription of target genes. Involved in the induction of several enzymes that participate in xenobiotic metabolism.
BATF::JUN	The protein encoded by this gene is a nuclear basic leucine zipper protein that belongs to the AP-1/ATF superfamily of transcription factors. The leucine zipper of this protein mediates dimerization with members of the Jun family of proteins. This protein is thought to be a negative regulator of AP-1/ATF transcriptional events.
Bhlhe40	This gene encodes a basic helix-loop-helix protein expressed in various tissues. The encoded protein can interact with ARNTL or compete for E-box binding sites in the promoter of PER1 and repress CLOCK/ ARNTL's transactivation of PER1. Transcriptional repressor involved in the regulation of the circadian rhythm by negatively regulating the activity of the clock genes and clock-controlled genes.

Appendix Continued

TFs	TF discription
BRCA1	This gene encodes a nuclear phosphoprotein that plays a role in maintaining genomic stability, and it also acts as a tumor suppressor.
CDX2	This gene is a member of the caudal-related homeobox transcription factor gene family. The encoded protein is a major regulator of intestine-specific genes involved in cell growth an differentiation. major regulator of intestine-specific genes involved in cell growth an differentiation.
CEBPA	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	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.
CRX	The protein encoded by this gene is a photoreceptor-specific transcription factor which plays a role in the differentiation of photoreceptor cells. This homeodomain protein is necessary for the maintenance of normal cone and rod function.
E2F6	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.
EBF1	Transcriptional activator which recognizes variations of the palindromic sequence 5'-ATTCCCNNGGGAATT-3'
ELK1	The protein encoded by this gene is a nuclear target for the ras-raf-MAPK signaling cascade.
EN1	Homeobox-containing genes are thought to have a role in controlling development.
FOS	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.
FOSL2	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. Activates CEBPB transcription in PGE2-activated osteoblasts.
FOXA1	Transcription factor that is involved in embryonic development, establishment of tissue-specific gene expression and regulation of gene expression in differentiated tissues.
FOXA2	Involved in embryonic development, establishment of tissue-specific gene expression and regulation of gene expression in differentiated tissues.
FOXC1	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.
FOXD1	This gene belongs to the forkhead family of transcription factors which are characterized by a distinct forkhead domain. Studies of the orthologous mouse protein indicate that it functions in kidney development by promoting nephron progenitor differentiation, and it also functions in the development of the retina and optic chiasm.

Appendix Continued

TFs	TF discription
FOXD3	This gene belongs to the forkhead family of transcription factors which is characterized by a distinct DNA-binding forkhead domain. Acts are a transcriptional activator and repressor.
FOXH1	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.
FOXI1	This gene belongs to the forkhead family of transcription factors which is characterized by a distinct DNA-binding forkhead domain. Transcriptional activator required for the development of normal hearing, sense of balance and kidney function.
FOXL1	FOX transcription factors are characterized by a distinct DNA-binding forkhead domain and play critical roles in the regulation of multiple processes including metabolism, cell proliferation and gene expression during ontogenesis. Transcriptional repressor. It plays an important role in the specification and differentiation of lung epithelium.
FOXO1	Transcription factor that is the main target of insulin signaling and regulates metabolic homeostasis in response to oxidative stress.
FOXO3	This gene belongs to the forkhead family of transcription factors which are characterized by a distinct forkhead domain. This gene likely functions as a trigger for apoptosis through expression of genes necessary for cell death.
FOXP1	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. Plays an important role in the specification and differentiation of lung epithelium.
FOXQ1	This gene belongs to the forkhead family of transcription factors which is characterized by a distinct DNA-binding forkhead domain. Plays a role in hair follicle differentiation.
FOXP2	Transcriptional repressor that may play a role in the specification and differentiation of lung epithelium. May also play a role in developing neural, gastrointestinal and cardiovascular tissues.
GATA1	The protein plays an important role in erythroid development by regulating the switch of fetal hemoglobin to adult hemoglobin.
GATA2	A member of the GATA family of zinc-finger transcription factors that are named for the consensus nucleotide sequence they bind in the promoter regions of target genes and play an essential role in regulating transcription of genes involved in the development and proliferation of hematopoietic and endocrine cell lineages.
GATA3	Plays an important role in endothelial cell biology.
GATA4	This protein is thought to regulate genes involved in embryogenesis and in myocardial differentiation and function. Promotes cardiac myocyte enlargement.
HIF1a: ARNT	HIF1 is a homodimeric basic helix-loop-helix structure composed of HIF1a, the alpha subunit, and the aryl hydrocarbon receptor nuclear translocator (Arnt), the beta subunit. The protein encoded by HIF1 is a Per-Arnt-Sim (PAS) transcription factor found in mammalian cells growing at low oxygen concentrations. It plays an essential role in cellular and systemic responses to hypoxia.

TFs	TF discription
HLTF	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.
HNF1A	Transcriptional activator that regulates the tissue specific expression of multiple genes, especially in pancreatic islet cells and in liver.
HNF4a	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
HNF4g	Transcription factor. Has a lower transcription activation potential than HNF4-alpha
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.
JUN (var.2)	This gene is the putative transforming gene of avian sarcoma virus 17. It encodes a protein which is highly similar to the viral protein, and which interacts directly with specific target DNA sequences to regulate gene expression.
JUN::FOS	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.
JUNB	Transcription factor involved in regulating gene activity following the primary growth factor response. Binds to the DNA sequence 5-TGA[CG]TCA-3
KLF5	Transcription factor that binds to GC box promoter elements. Activates transcription of genes.
LHX3	This gene encodes a member a large protein family which carry the LIM domain, a unique cysteine-rich zinc-binding domain. The encoded protein is a transcription factor that is required for pituitary development and motor neuron specification.
MAFB	The encoded nuclear protein represses ETS1-mediated transcription of erythroid-specific genes in myeloid cells. This protein plays an essential role in the regulation of hematopoiesis and may play a role in tumorigenesis.
MAFF	The protein encoded by this gene is a basic leucine zipper (bZIP) transcription factor that lacks a transactivation domain. Interacts with the upstream promoter region of the oxytocin receptor gene.
MAFK	Since they lack a putative transactivation domain, the small Mafs behave as transcriptional repressors when they dimerize among themselves. they seem to serve as transcriptional activators by dimerizing with other (usually larger) basic-zipper proteins and recruiting them to specific DNA-binding sites. Small Maf proteinS heterodimerize with Fos and may act as competitive repressors of the NF-E2 transcription factor.
MAX	The protein encoded by this gene is a member of the basic helix-loop-helix leucine zipper (bHLHZ) family of transcription factors
MEF2C	Transcription activator which binds specifically to the MEF2 element present in the regulatory regions of development. many muscle-specific genes. Controls cardiac morphogenesis and myogenesis, and is also involved in vascular development.

Appendix Continued

TFs	TF discription
MZF1_1-4	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.
MZF1_5-13	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.
NFE2L1:MAFG	Nuclear factor erythroid 2-related factor (Nrf2) coordinates the up-regulation of cytoprotective genes via the antioxidant response element (ARE). MafG is a ubiquitously expressed small maf protein that is nvolved in cell differentiation of erythrocytes. It dimerizes with P45 NF-E2 protein and activates expression of a and b-globin.
NFIL3	Expression of interleukin-3 (IL3; MIM 147740) is restricted to activated T cells, natural killer (NK) cells, and mast cell lines.
NKX2-5	This gene encodes a member of the NK family of homeobox-containing proteins. Transcriptional repressor that acts as a negative regulator of chondrocyte maturation.
NKX3-1	This gene encodes a homeobox-containing transcription factor. This transcription factor functions as a negative regulator of epithelial cell growth in prostate tissue.
Nr1h3::Rxra	The protein encoded by this gene belongs to the NR1 subfamily of the nuclear receptor superfamily. The NR1 family members are key regulators of macrophage function, controlling transcriptional programs involved in lipid homeostasis and inflammation. This protein is highly expressed in visceral organs, including liver, kidney and intestine. It forms a heterodimer with retinoid X receptor (RXR), and regulates expression of target genes containing retinoid response elements. Studies in mice lacking this gene suggest that it may play an important role in the regulation of cholesterol homeostasis.
NR3C1	Glucocorticoids regulate carbohydrate, protein and fat metabolism, modulate immune responses through supression of chemokine and cytokine production and have critical roles in constitutive activity of the CNS, digestive, hematopoietic, renal and reproductive systems.
PAX2	Probable transcription factor that may have a role in kidney cell differentiation.
PDX1	Activates insulin, somatostatin, glucokinase, islet amyloid polypeptide and glucose transporter type 2 gene transcription. Particularly involved in glucose-dependent regulation of insulin gene transcription.
POU5F1::SOX2	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.
PRRX2	The DNA-associated protein encoded by this gene is a member of the paired family of homeobox proteins. Expression is localized to proliferating fetal fibroblasts and the developing dermal layer, with downregulated expression in adult skin.
RFX1	This gene is a member of the regulatory factor X gene family, which encodes transcription factors that contain a highly-conserved winged helix DNA binding domain. The protein encoded by this gene is structurally related to regulatory factors X2, X3, X4, and X5. Regulatory factor essential for MHC class II genes expression. Binds to the X boxes of MHC class II genes.

TFs	TF discription
RFX5	Activates transcription from class II MHC promoters. Recognizes X-boxes.
RORA_1	The protein encoded by this gene is a member of the NR1 subfamily of nuclear hormone receptors. 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.
RUNX2	Transcription factor involved in osteoblastic differentiation and skeletal morphogenesis. Essential for the maturation of osteoblasts and both intramembranous and endochondral ossification.
RXRa	Retinoid X receptors (RXRs) and retinoic acid receptors (RARs), are nuclear receptors that mediate the biological effects of retinoids by their involvement in retinoic acid-mediated gene activation.
SOX2	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. The product of this gene is required for stem-cell maintenance in the central nervous system, and also regulates gene expression in the stomach.
SOX3	Transcription factor required during the formation of the hypothalamo-pituitary axis. May function as a
	switch in neuronal development. Keeps neural cells undifferentiated by counteracting the activity of proneural
SOX5	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.
	The encoded protein may act as a transcriptional regulator after forming a protein complex with other proteins
	and may play a role in chondrogenesis.
SOX6	The encoded protein is a transcriptional activator that is required for normal
	development of the central nervous system, chondrogenesis and maintenance of cardiac and skeletal muscle cells.
SOX9	Plays an important role in the normal skeletal development. May regulate the expression of other genes
	involved in chondrogenesis by acting as a transcription factor for these genes
SOX10	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.
SOX17	Acts as transcription regulator that binds target promoter DNA and bends the DNA.
SP1	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.
SPI1	This gene encodes an ETS-domain transcription factor that activates gene expression during myeloid
	and B-lymphoid cell development
SPIB	The protein encoded by this gene is a transcriptional activator that binds to the
	PU-box (5'-GAGGAA-3') and acts as a lymphoid-specific enhancer.
SRF	This gene encodes a ubiquitous nuclear protein that stimulates both cell proliferation and differentiation.
	This protein binds to the serum response element (SRE) in the promoter region of target genes.
	Required for cardiac differentiation and maturation.

Appendix Continued

TFs	TF discription
SREBF1	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	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	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
STAT1	The protein encoded by this gene is a member of the STAT protein family. In response to cytokines and growth factors, STAT family members are phosphorylated by the receptor associated kinases, and then form homo- or heterodimers that translocate to the cell nucleus where they act as transcription activators. This protein can be activated by various ligands including interferon-alpha, interferon-gamma, EGF, PDGF and IL6. This protein mediates the expression of a variety of genes, which is thought to be important for cell viability in response to different cell stimuli and pathogens.
STAT3	Signal transducer and transcription activator that mediates cellular responses to interleukins, KITLG/SCF and other growth factors
STAT4	Carries out a dual function: signal transduction and activation of transcription. Involved in IL12 signaling This protein is essential for mediating responses to IL12 in lymphocytes, and regulating the differentiation of arthritis. T helper cells. Mutations in this gene may be associated with systemic lupus erythematosus and rheumatoid arthritis.
STAT5A:STAT5B	Carries out a dual function: signal transduction and activation of transcription. Regulates the expression of milk proteins during lactation.
STAT6	This protein plays a central role in exerting IL4 mediated biological responses. It is found to induce the expression of BCL2L1/BCL-X(L), which is responsible for the anti-apoptotic activity of IL4. Carries out a dual function: signal t ransduction and activation of transcription. Involved in IL4/interleukin-4- and IL3/interleukin-3-mediated signaling.
T	The protein encoded by this gene is an embryonic nuclear transcription factor that binds to a specific DNA element, the palindromic T-site. It binds through a region in its N-terminus, called the T-box, and effects transcription of genes required for mesoderm formation and differentiation.
TBP	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.
TFAP2C	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.
THAP1	DNA-binding transcription regulator that regulates endothelial cell proliferation and G1/S cell-cycle progression.
ZBTB33	This gene encodes a transcriptional regulator with bimodal DNA-binding specificity, which binds to methylated CGCG and also to the non-methylated consensus KAISO-binding site TCCTGCNA. The protein contains an N-terminal POZ/BTB domain and 3 C-terminal zinc finger motifs. It recruits the N-CoR repressor complex to promote histone deacetylation and the formation of repressive chromatin structures in target gene promoters. It may contribute to the repression of target genes of the Wnt signaling pathway, and may also activate transcription of a subset of target genes by the recruitment of catenin delta-2 (CTNND2).
ZNF263	Might play an important role in basic cellular processes as a transcriptional repressor.
ZNF354C	May function as a transcription repressor.

epithelium, negative regulation of chondrocyte maturation, regulation of cholesterol homeostasis, glucose-dependent regulation of insulin gene transcription, proliferating fetal fibroblasts and the developing dermal layer, embryonic development and male development, respectively (Figure 1, Table 2, Appendix). There are also six conserved TBFS for the EN1, FOXA1, FOXA2, HLTF, RUNX2 and SOX17 TFs which are involved controlling development, embryonic development, altering chromatin structure, osteoblastic differentiation and transcription regulation, respectively (Table 2, Appendix).

The common rs11889341 SNP *STAT4*-C allele creates three unique TBFS for the AR, CEBPb and the SOX3 TFs, which are involved with steroid-hormone activation, the regulation of acute-phase reaction, inflammation and hemopoiesis and the formation of the hypothalamo-pituitary axis, respectively (Figure 2, Table 2, Appendix). The minor *STAT4*-T allele creates thirteen unique TFBS for the CDX2, FOXD1, FOXD3, FOXI1, FOXO1, FOXO3, FOXP1, FOXPW, GATA1-4 and MEF2C TFs which are involved with the regulation of intestine-specific genes, kidney development, transcriptional activation and repression, kidney function, metabolic homeostasis in response to oxidative stress, a trigger for apoptosis, differentiation of lung epithelium, differentiation of lung epithelium, switching from fetal to adult hemoglobin, development and proliferation of hematopoietic and endocrine cell lineages, endothelial cell biology, cardiac myocyte enlargement and vascular development, respectively (Table 2, Appendix). There are also seven conserved TFBS for FOXL1, HLTF, NR3C1, SOX5, SOX6, SRY and ZNF263 TFs which are involved in the specification and differentiation of lung epithelium, altering chromatin structure, modulation of immune responses through suppression of chemokine and cytokine production, regulation of embryonic development, development of the central nervous system, a genetic switching in male development and transcription repression, respectively (Table 2, Appendix).

The common rs8179673 SNP *STAT4*-T allele creates nine unique TFBS for the ARID3A, EN1, FOXL1, GATA3, HOXA5, LHX3, NFIL3, NKX3-1 and RORa_1 TFs which are involved with involved with the control of cell cycle progression, the specification and differentiation of lung epithelium, endothelial cell biology, developmental regulatory system, pituitary development and motor neuron specification, expression of interleukin-3, regulation of embryonic development, cellular

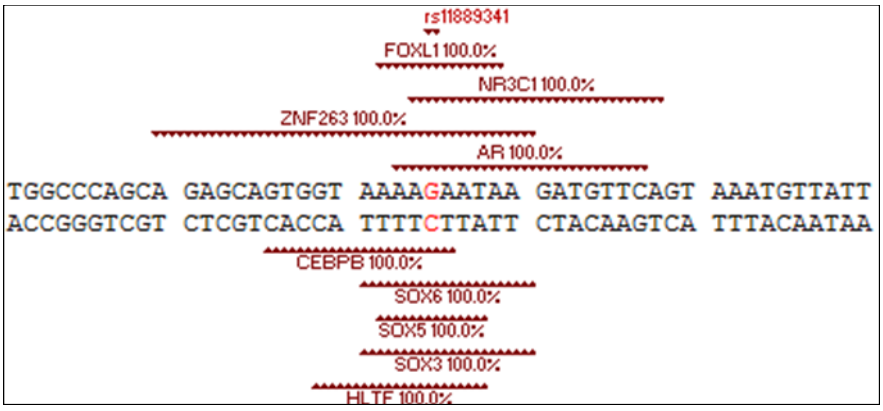

Figure 2. Double stranded DNA from the *STAT4* gene showing the potential TFBS for nine different TFs which can bind their respective DNA sequence either above (+) or below (-) the duplex (cf. Table 2). The rs11889341 rSNP common *STAT4*-C allele is found in each of these TFBS. As shown, this rSNP is located in the 70 kb intron between exon 2 and 3 of the *STAT4* gene. Also included with the potential TFBS is their % sequence homology to the duplex.

differentiation, immunity, circadian rhythm as well as lipid, steroid, xenobiotics and glucose metabolism, respectively (Table 2, Appendix). The minor rs8179673 SNP *STAT4*-C allele creates fifteen unique TFBS for the BRCA1, FOXA2, FOXD1, FOXO1, FOXP1 & 2, FOXQ1, HLTF, HNF1a, HNF4a, HNF4g, JUN::FOX, SOX5 and SOX6 TFs which are involved with tumor suppression, embryonic development, kidney development, insulin signaling, differentiation of lung epithelium, hair follicle differentiation, altering chromatin structure, regulation of the tissue specific expression of pancreatic islet cells and liver, regulation of several hepatic genes, cell proliferation and differentiation, transcriptional regulation and activation, respectively (Table 2, Appendix). There are also eight conserved TFBS for the FOXA1, RXRa, SOX2, SOX3, SOX6, SOX10, SRY and TBP TFs which are involved with embryonic development, retinoic acid-mediated gene activation, regulation of embryonic development, the formation of the hypothalamo-pituitary axis, development of the central nervous system, a genetic switching in male development and binding of TFIID to the TATA box, respectively (Table 2, Appendix).

The common rs7582694 SNP *STAT4*-G allele creates four unique TFBS for the MAX, SOX3, SOX10 and T TFs which are involved with transcription regulation, the formation of the hypothalamo-pituitary axis, regulation of embryonic development and mesoderm formation and differentiation, respectively (Table 2, Appendix). The minor *STAT4*-C allele creates three unique TFBS for the BH1HE40, HIF1a::ARNT and SRY TFs which are involved with the regulation of circadian rhythm, cellular and systemic responses to hypoxia, and a genetic switch in male development, respectively (Table 2, Appendix). There are also eight conserved TFBS for the BATF::JUN, CEBPa, FOSL2, FOXC1, HLTF, JUN (var.2) and JUNB TFs which are involved in negative regulation of AP-1/ATF transcriptional events, cell cycle regulation and body weight homeostasis, regulation of cell proliferation, differentiation, and transformation, cell viability and resistance to oxidative stress, altering chromatin structure, regulation of gene expression and gene activity, respectively (Table 2, Appendix).

The common rs7574070 SNP *STAT4*-A allele creates seven unique TFBS for the CEBPb, MZF1_1-4, NFEL1::MAFG, PAX2, RFX1 and RFX5 TFs which are involved with the regulation of acute-phase reaction, inflammation and hemopoiesis, hemopoietic development, up-regulation of cytoprotective genes, kidney cell differentiation and the activation of transcription from class II MHC promoters, respectively (Table 2, Appendix). The minor rs7574070 SNP *STAT4*-C allele creates five unique TFBS for the EBF1, ELK1, STAT1, STAT4 and TFAP2C TFs with are involved with transcriptional activation, the ras-raf-MAPK signaling cascade, transcriptional activation for cell viability in response to different cell stimuli and pathogens and activation of genes involved in a large spectrum of biological developmental functions, respectively (Table 2, Appendix). There are also three conserved TFBS for the STAT3, STAT5A::STAT5B and THAP1 TFs which are involved with signal transduction and transcriptional activation as well as the regulation of endothelial cell proliferation and the G1/S cell-cycle, respectively (Table 2, Appendix).

The common rs7572482 SNP *STAT4*-A allele creates eight unique TFBS for the ARID3A, EBF1, FOS, NFE2L1::MAFG, POU5F1::SOX2, RFX1, SPIB and STAT5A::STAT5B TFs which are involved with the control of cell cycle progression, transcriptional activation, regulation of cell proliferation, differentiation, and transformation, up-regulation of cytoprotective genes, embryonic development and stem cell pluripotency, regulation factor essential for MHC class II genes expression, lymphoid-specific enhancer, signal transduction and activation of transcription, respectively (Table 2, Appendix). The minor common rs7572482 SNP

STAT4-G allele creates four unique TFBS for the ARNT, ARNT::AHR, SPI1 and ZBTB33 TFs which are involved with xenobiotic metabolism, activates gene expression during myeloid and B-lymphoid cell development, and transcriptional regulation with bimodal DNA-binding specificity, respectively (Table 2, Appendix). There are also four conserved TFBS for the SOX3, STAT3, STAT4 and STAT6 TFs which are involved with the formation of the hypothalamo-pituitary axis, signal transduction and transcriptional activation, mediating responses to IL12 in lymphocytes, and regulating the differentiation of arthritis, exerting IL4 mediated biological responses, respectively (Table 2, Appendix).

The remaining two *STAT4* SNPs (rs7568275 and rs10181656) that have been found to be significantly associated with human disease (Table 1) can be analyzed in a similar fashion as the SNPs above (Table 2, Appendix).

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 [48, 49] and the remaining 93% are located within non-coding areas [50, 51] 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 [25, 27, 52]. 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 [52]. Examples of rSNPs associated with disease susceptibility are numerous and several reviews have been published.[52-55].

The rs7574865 rSNP *STAT4*-G allele [G (+ strand) or C (- strand)] located in the unique MAX and ZNF354C TFBS have a 100% occurrence in humans while the unique FOXL1 TFBS has a 17% occurrence (Table 2). Since these binding sites (BS) occur multiple times in the gene, the rSNP G allele should not have much of an impact gene regulation (Table 2). The minor rs7574865 rSNP *STAT4*-T allele [T (+ strand) or A (- strand)] located in the unique ARID3A, FOXQ1 and PDX1 TFBS have a 100% occurrence in humans while the NR1HE::RXRa TFBS has a 44% occurrence (Figure 1, Table 2). Since all the unique BS for this allele occur multiple times in this gene, it would not be expected that these TFBS would have much of an effect on *STAT4* regulation except for the NR1HE:: RXRa BS which occurs only once in the gene (Table 2) and is a key regulator of macrophage function, controlling transcriptional programs involved in lipid homeostasis and inflammation (Appendix). Since NR1HE:: RXRa protein duplex is part of the NR1 subfamily of the retinoid nuclear receptor superfamily, the presence of its TFBS created by the minor T allele could in part be responsible for the diseases listed in Table 1, that are significantly associated with this rSNP.

The rs11889341 rSNP *STAT4*-C allele [C (+ strand) or G (- strand) located in the unique AR TFBS has a 100% occurrence in humans and occurs only once in the gene (Figure 2, Table 2). The androgen receptor is a steroid-hormone activated transcription factor which stimulates transcription of androgen responsive genes that are expressed in bone marrow, mammary gland, prostate, testicular and muscle tissues. The absence of

this TFBS created by the minor *STAT4*-T allele should have a major effect relating to the diseases listed in Table 1. The minor rs11889341 rSNP *STAT4*-T allele [T (+ strand) or A (- strand)] located in the unique CDX2, FOXOD1, FOXI1, FOXO1, FOXP1, FOXP2 and GATA3 TFBS have a 100% occurrence in humans. These TFBS occur multiple times in the gene so they would not be expected to have much impact on the regulation of the gene, except for the FOXP1 TFBS which occurs only once in the gene (Table 2). Although FOXP1 is a member of the subfamily P of the forkhead box (FOX) transcription factors which play important roles in the regulation of tissue- and cell type-specific gene transcription during both development and adulthood, it is doubtful that the presence of this TFBS only with the minor T allele would have much impact on the regulation of the gene since there are other family members TFBS represented with the minor allele (Table 2). The SNP T allele is also located in the two unique MEF2C TFBS which have a 95 and 97% occurrence in humans and each occurs only once in the gene (Table 2). The MEF2C TF controls cardiac morphogenesis and myogenesis, and is also involved in vascular development and consequently the presence or absence of this TFBS should have an impact on the diseases listed in Table 1.

The rs8179673 rSNP *STAT4*-T allele [T (+ strand) or A (- strand)] located in the unique ARID3A, GATA3 and NKX3-1 TFBSs have a 100% occurrence in humans while the LHX3 and NFIL3 TFBS have a 95 and 96% occurrence (Table 2). Since these TFBS occur more than once in the gene, it is doubtful that these BS would have much impact on the regulation of the gene. The minor rs8179673 rSNP *STAT4*-C allele [C (+ strand) or G (- strand)] located in the unique FOXH1, FOXO1, FOXQ1, SOX2 and SOX6 TFBS have a 100% occurrence in humans while the HNF1a and HNF4g TFBS have a 67 and 93% occurrence, respectively (Table 2). All of these TFBS occur only once in the gene (Table 2). However, the FOXH1, FOXO1, FOXQ1, SOX2 and SOX6 TFBS are involved with transcription machinery and are represented in families consisting of multiple members. Therefore, it is unlikely that the presence or absence of one family member would have much impact on gene regulation especially since there are TFBS for multiple family members represented by the minor C allele (Table 2, Appendix). The unique HNF1a and HNF4g TFBS which also occur only once in the gene are BS for the HNF1a and HNF4g transcriptional activators that regulates the tissue specific expression of multiple genes, especially in pancreatic islet cells and in liver (Table 2, Appendix).

The rs7574070 rSNP *STAT4*-A allele [A (+ strand) or T (- strand)] located in the unique CEBPb and ZNF354C TFBS have a 100% occurrence in humans while the RFX1 and RFX5 TFBS have a 84 and 78% occurrence, respectively (Table 2). The CEBPb, RFX1 and RFX5 TFBS occur only once in the gene while the ZNF354C TFBS occurs 87 times in the gene (Table 2); consequently, only the CEBPb, RFX1 and RFX5 TFBS might have an impact on gene regulation. The CEBPb TF is an important transcriptional activator regulating the expression of genes involved in immune and inflammatory responses; consequently, the loss of the BS with the minor C allele could have an impact on Behcet's disease (Table 1). The RFX1 & 5 TFs are important regulatory factors essential for MHC class II gene expression and the loss of these BS with the presence of the minor C allele could also have an impact on Behcet's disease (Table 1). The rs7574070 rSNP *STAT4*-C allele [C (+ strand) or G (- strand)] located in the unique STAT4 and THAP2C TFBS have a 43 and 99% occurrence in humans and occur only once in the gene (Table 2). The STAT4 TF is important in regulating genes associated with systemic lupus erythematosus and rheumatoid arthritis (Appendix); consequently, the occurrence of this TFBS only with the minor C allele could have an impact on Behcet's disease (Table 1). The THAP2C TF is involved with a large spectrum of

important biological functions (Appendix) and also contribute to Behcet's disease when the TFBS is only represented in the minor C allele (Table 2, Appendix).

The rs7572482 rSNP *STAT4*-A allele [A (+ strand) or T (- strand)] located in the unique ARID3A, FOS, NFE2L1::MAFG and POU5F1::SOX2 TFBS have a 100% occurrence in humans except for the POU5F1::SOX2 TFBS for which it has a 90% occurrence (Table 2). The ARID3A and NFE2L1::MAFG TFBS occur multiple times in the gene while the FOS and POU5F1::SOX2 TFBS only occurs once (Table 2). The FOS TF is a regulator of cell proliferation, differentiation and transformation while the POU5F1::SOX2 TFs play a key role in embryonic development and stem cell pluripotency (Appendix) which could have an impact on Behcet's disease (Table 1). The rs7572482 rSNP *STAT4*-G allele [G (+ strand) or C (- strand)] located in the unique ARNT, ARNT::AHR and ZBTB33 TFBS have a 100% occurrence in humans except for the ARNT::AHR TFBS for which it has a 96% occurrence (Table 2). The ARNT and ARNT::AHR TFBS occur 14 times in the *STAT4* gene while the ZBTB33 TFBS only occurs once. The ZBTB33 TF is a transcriptional regulator involved with zinc finger motifs and may contribute to the repression of target genes of the Wnt signaling pathway (Appendix). Similar logic can be used to evaluate the potential TFBS within the other *STAT4* rSNPs found in the Tables 1 & 2.

Human diseases or conditions can be associated with rSNPs of the *STAT4* 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 potential TFBS associated with the rs7574865 rSNP *STAT4*-T allele from Table 2 are illustrated in Figure 1 as well as the rs11889341 common rSNP *STAT4*-C allele illustrated in Figure 2. As can be seen in Table 2, these potential TFBS change when an individual carries the alternate allele. The importance of this has been illustrated in Figure 2 with the AR TFBS where the common C allele has this function and the minor T allele does not. The AR TF is a steroid-hormone activated transcription factor which stimulates transcription of androgen responsive genes. A second example would be the HNF1a and HNF4g TFBS where the minor rs8179673 T allele has this function while the common allele does not. This TF regulates the tissue specific expression of multiple genes, especially in pancreatic islet and liver cells. A third example found with the rs7574070 common rSNP *STAT4*-A allele and not for the minor C allele is for the RFX1 & 5 TFBS whose TFs are important regulatory factors essential for MHC class II gene expression. Other examples can be found in Table 2.

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 *STAT4* gene can provide different TFBS which in turn are also associated with disease in humans. 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 INTERESTS

Author has declared that no competing interests exist.

APPENDIX

Supplemental material is available for this article.

REFERENCES

1. Wang Y, Qu A, Wang H, Signal transducer and activator of transcription 4 in liver diseases. Int J Biol Sci;11: 448-55.
2. Khanna P, Chua PJ, Bay BH, Baeg GH, The JAK/STAT signaling cascade in gastric carcinoma (Review). Int J Oncol.
3. Heneghan AF, Pierre JF, Kudsk KA, JAK-STAT and intestinal mucosal immunology. JAKSTAT;2: e25530.
4. O'Shea JJ, Schwartz DM, Villarino AV, Gadina M, McInnes IB, Laurence A, The JAK-STAT pathway: impact on human disease and therapeutic intervention. Annu Rev Med;66: 311-28.
5. Taylor KE, Remmers EF, Lee AT, Ortmann WA, Plenge RM, Tian C, Chung SA, Nititham J, Hom G, Kao AH, Demirci FY, Kamboh MI, Petri M, Manzi S, Kastner DL, Seldin MF, Gregersen PK, Behrens TW, Criswell LA, Specificity of the STAT4 genetic association for severe disease manifestations of systemic lupus erythematosus. PLoS Genet 2008;4: e1000084.
6. Jiang DK, Sun J, Cao G, Liu Y, Lin D, Gao YZ, Ren WH, Long XD, Zhang H, Ma XP, Wang Z, Jiang W, Chen TY, Gao Y, Sun LD, Long JR, Huang HX, Wang D, Yu H, Zhang P, Tang LS, Peng B, Cai H, Liu TT, Zhou P, Liu F, Lin X, Tao S, Wan B, Sai-Yin HX, Qin LX, Yin J, Liu L, Wu C, Pei Y, Zhou YF, Zhai Y, Lu PX, Tan A, Zuo XB, Fan J, Chang J, Gu X, Wang NJ, Li Y, Liu YK, Zhai K, Hu Z, Liu J, Yi Q, Xiang Y, Shi R, Ding Q, Zheng W, Shu XO, Mo Z, Shugart YY, Zhang XJ, Zhou G, Shen H, Zheng SL, Xu J, Yu L, Genetic variants in STAT4 and HLA-DQ genes confer risk of hepatitis B virus-related hepatocellular carcinoma. Nat Genet;45: 72-5.
7. Hou S, Yang Z, Du L, Jiang Z, Shu Q, Chen Y, Li F, Zhou Q, Ohno S, Chen R, Kijlstra A, Rosenbaum JT, Yang P, Identification of a susceptibility locus in STAT4 for Behcet's disease in Han Chinese in a genome-wide association study. Arthritis Rheum;64: 4104-13.
8. Bolin K, Sandling JK, Zickert A, Jonsen A, Sjowall C, Svenungsson E, Bengtsson AA, Eloranta ML, Ronnblom L, Syvanen AC, Gunnarsson I, Nordmark G, Association of STAT4 polymorphism with severe renal insufficiency in lupus nephritis. PLoS One;8: e84450.
9. Kim ES, Kim SW, Moon CM, Park JJ, Kim TI, Kim WH, Cheon JH, Interactions between IL17A, IL23R, and STAT4 polymorphisms confer susceptibility to intestinal Behcet's disease in Korean population. Life Sci;90: 740-6.
10. Lu Y, Zhu Y, Peng J, Wang X, Wang F, Sun Z, STAT4 genetic polymorphisms association with spontaneous clearance of hepatitis B virus infection. Immunol Res;62: 146-52.
11. Yi J, Fang X, Wan Y, Wei J, Huang J, STAT4 polymorphisms and diabetes risk: a meta-analysis with 18931 patients and 23833 controls. Int J Clin Exp Med;8: 3566-72.
12. Kumar A, Das S, Agrawal A, Mukhopadhyay I, Ghosh B, Genetic association of key Th1/Th2 pathway candidate genes, IRF2, IL6, IFNGR2, STAT4 and IL4RA, with atopic asthma in the Indian population. J Hum Genet;60: 443-8.
13. Mathur AN, Chang HC, Zisoulis DG, Stritesky GL, Yu Q, O'Malley JT, Kapur R, Levy DE, Kansas GS, Kaplan MH, Stat3 and Stat4 direct development of IL-17-secreting Th cells. J Immunol 2007;178: 4901-7.

14. Chitnis T, Najafian N, Benou C, Salama AD, Grusby MJ, Sayegh MH, Khoury SJ, Effect of targeted disruption of STAT4 and STAT6 on the induction of experimental autoimmune encephalomyelitis. J Clin Invest 2001;108: 739-47.

15. Mo C, Chearwae W, O'Malley JT, Adams SM, Kanakasabai S, Walline CC, Stritesky GL, Good SR, Perumal NB, Kaplan MH, Bright JJ, Stat4 isoforms differentially regulate inflammation and demyelination in experimental allergic encephalomyelitis. J Immunol 2008;181: 5681-90.

16. Remmers EF, Plenge RM, Lee AT, Graham RR, Hom G, Behrens TW, de Bakker PI, Le JM, Lee HS, Batliwalla F, Li W, Masters SL, Booty MG, Carulli JP, Padyukov L, Alfredsson L, Klareskog L, Chen WV, Amos CI, Criswell LA, Seldin MF, Kastner DL, Gregersen PK, STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. N Engl J Med 2007;357: 977-86.

17. Mudter J, Weigmann B, Bartsch B, Kiesslich R, Strand D, Galle PR, Lehr HA, Schmidt J, Neurath MF, Activation pattern of signal transducers and activators of transcription (STAT) factors in inflammatory bowel diseases. Am J Gastroenterol 2005;100: 64-72.

18. Hou S, Kijlstra A, Yang P, The genetics of Behcet's disease in a Chinese population. Front Med;6: 354-9.

19. Liao Y, Cai B, Li Y, Chen J, Ying B, Tao C, Zhao M, Ba Z, Zhang Z, Wang L, Association of HLA-DP/DQ, STAT4 and IL-28B variants with HBV viral clearance in Tibetans and Uygurs in China. Liver Int;35: 886-96.

20. Kim LH, Cheong HS, Namgoong S, Kim JO, Kim JH, Park BL, Cho SW, Park NH, Cheong JY, Koh I, Shin HD, Kim YJ, Replication of genome wide association studies on hepatocellular carcinoma susceptibility loci of STAT4 and HLA-DQ in a Korean population. Infect Genet Evol;33: 72-6.

21. Liu QF, Li Y, Zhao QH, Wang ZY, Hu S, Yang CQ, Ye K, Li L, Association of STAT4 rs7574865 polymorphism with susceptibility to inflammatory bowel disease: A systematic review and meta-analysis. Clin Res Hepatol Gastroenterol.

22. Fan ZD, Wang FF, Huang H, Huang N, Ma HH, Guo YH, Zhang YY, Qian XQ, Yu HG, STAT4 rs7574865 G/T and PTPN22 rs2488457 G/C polymorphisms influence the risk of developing juvenile idiopathic arthritis in Han Chinese patients. PLoS One;10: e0117389.

23. Aiba Y, Yamazaki K, Nishida N, Kawashima M, Hitomi Y, Nakamura H, Komori A, Fuyuno Y, Takahashi A, Kawaguchi T, Takazoe M, Suzuki Y, Motoya S, Matsui T, Esaki M, Matsumoto T, Kubo M, Tokunaga K, Nakamura M, Disease susceptibility genes shared by primary biliary cirrhosis and Crohn's disease in the Japanese population. J Hum Genet.

24. Xu L, Dai WQ, Wang F, He L, Zhou YQ, Lu J, Xu XF, Guo CY, Association of STAT4 gene rs7574865G > T polymorphism with ulcerative colitis risk: evidence from 1532 cases and 3786 controls. Arch Med Sci;10: 419-24.

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

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

27. 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.

28. 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.

29. 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.
30. 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.
31. 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.
32. 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.
33. 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.
34. 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.
35. 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.
36. 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.
37. 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.
38. Kumar A, Purohit R, Computational investigation of pathogenic nsSNPs in CEP63 protein. Gene 2012;503: 75-82.
39. Kamaraj B, Purohit R, Computational screening of disease-associated mutations in OCA2 gene. Cell Biochem Biophys 2014;68: 97-109.
40. 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.
41. Kumar A, Purohit R, Use of long term molecular dynamics simulation in predicting cancer associated SNPs. PLoS Comput Biol 2014;10: e1003318.
42. Liu L, Zhao W, Zhou X, Modeling co-occupancy of transcription factors using chromatin features. Nucleic Acids Res.
43. Liu L, Jin G, Zhou X, Modeling the relationship of epigenetic modifications to transcription factor binding. Nucleic Acids Res;43: 3873-85.
44. 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.

45. 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.

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

47. 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.

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

49. 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.

50. 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.

51. 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.

52. 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.

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

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

55. 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.