Open Journal of Genetics, 2013, 3, 253-261 **OJGen**
http://dx.doi.org/10.4236/ojgen.2013.34028 Published Online December 2013 (http://www.scirp.org/journal/ojgen/)

ATF3 rSNPs, transcriptional factor binding sites and human etiology*

Norman E. Buroker

Department of Pediatrics, University of Washington, Seattle, USA
Email: nburoker@u.washington.edu

Received 26 September 2013; revised 25 October 2013; accepted 15 November 2013

ABSTRACT

Three rSNPs (rs3125289, rs1877474 and rs11119982) in one intron of the activating transcription factor 3 (*ATF*3) gene have been significantly associated with the human etiology of hypospadias and may be associated with human disease. These rSNP alleles alter the DNA landscape for potential transcriptional factors (TFs) to attach resulting in changes in transcriptional factor binding sites (TFBS). These TFBS changes are examined with respect to the human etiology of hypospadias which has been found to be significantly associated with the rSNPs.

Keywords: *ATF*3; rSNPs; TFBS; Human Etiology

1. INTRODUCTION

The activating transcription factor 3 (*ATF*3) gene is a member of the activating transcription factor/cAMP responsive element binding (CREB) protein family of transcription factors, which share the basic region-leucine zipper (bZip) DNA binding motif (TGACGTCA). This gene is induced by a variety of signals including many of those encountered by cancer cells, and is involved in the complex process of cellular stress response [1-3]. *ATF*3 has been viewed as a hub of the cellular adaptive response network which allows cells adapt to disturbances in homeostasis [4]. This gene has been shown to be up-regulated during sexual differentiation [5] which indicates a potential role in hypospadias. Three unlinked *ATF*3 SNPs (rs3125289, rs1877474 and rs11119982) which span a 16 kb region of intron one have been independently found to be significantly associated with the risk of hypospadias [6]. This suggests that these SNPs in intron one may be part of a regulatory region for the *ATF*3 gene. To follow up on this possibility, the SNPs were examined for associations to potential transcription factor binding sites (TFBS).

Single nucleotide changes that affect gene expression by impacting gene regulatory sequences such as promoters, enhancers, and silencers are known as regulatory SNPs (rSNPs) [7-10]. A rSNPs within a transcriptional factor binding site (TFBS) can change a transcriptional factor's (TF) ability to bind its TFBS [11-14], in which case the TF would be unable to effectively regulate its target gene [15-19]. This concept is examined for the above *ATF*3 SNPs of intron one and their allelic association with TFBS. In this report, I discuss these SNP associations with changes in potential TFBS and their possible relationship to human etiology.

2. MATERIALS AND METHODS

Identifying TFBS

The JASPAR CORE database [20,21] and ConSite [22] were used to identify the TFBS in this study. JASPAR is a collection of transcription factor DNA-binding preferencesused for scanning genomic sequences where ConSite is a web-based tool for finding cis-regulatory elements in genomic sequences. The Vector NTI Advance 11 computer program (Invitrogen, Life Technologies) was used to locate the TFBS in the *ATF*3 gene (NCBI Ref Seq NM_001674) from 51 kb upstream of the transcriptional start site to 766 bp past the 3'UTR which represents a total of 57.6 Kbp. The JASPAR CORE database was also used to compute each nucleotide occurrence (%) within the TFBS where upper case lettering indicate that the nucleotide occurs 90% or greater and lower case less than 90%. The occurrence of each SNP allele in the TFBS is also computed from the database (**Table 1** & **Supplement**).

3. RESULTS

ATF3 rSNPs and TFBS

The *ATF*3 gene encodes a member of the mammalian

*Conflict of interest statement: The author declares that there is no conflict of interests.

Table 1. The ATF3 SNPs from intron 1 that were examined in this study. 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 rSNP 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 numbers (#) 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
rs3125289	**C**	ARNT	aryl hydrocarbon receptor nuclear translocator	7	cAcGaG	minus
(**C**/T)					G = 100%	
		ARNT	aryl hydrocarbon receptor nuclear translocator	7	ctCGTG	plus
					C = 100%	
		ARNT:AHR	aryl hydrocarbon receptor nuclear translocator	7	CtCGTG	plus
			aryl hydrocarbon receptor		C = 96%	
		ELF5	E74-like factor 5	17	agctTCCtc	plus
			(ets domain transcription factor)		c = 18%	
		FOXC1	Forkhead box C1	1	ctcgtGTA	plus
					c = 31%	
		GABPA	GA binding protein transcription factor, alpha	1	gaGGAAgctct	minus
					g = 19%	
		MYC	v-*myc* myelocytomatosis viral oncogene homolog	1	taCACGaGga	minus
					G = 90%	
		MYCN	v-myc myelocytomatosis viral related	1	taCACGaGga	minus
			oncogene, neuroblastoma derived		G = 90%	
		MZF1_1-4	Myeloid zinc finger 1	11	cGaGGA	minus
					G = 95%	
		SPIB	Spi-B transcription factor (Spi-1/PU.1 related)	12	cgaGGAA	minus
					g = 59%	
		TBP	TATA box binding protein	1	ctATAtAcacgagga	minus
					g = 33%	
		USF1	Upstream transcription factor 1	2	CtCGTGt	plus
					C = 93%	
		USF1	Upstream transcription factor 1	3	CACGaGg	minus
					G = 97%	
	T	ELF5	E74-like factor 5	2	agctTCCtt	plus
			(ets domain transcription factor)		t = 48%	
		FOXA1	Forkhead box A1	1	TgTgTatatag	plus
					T = 98%	
		FOXA2	Forkhead box A2	1	TgTgtatatagg	plus
					T = 100%	
		FOXA2	Forkhead box A2	1	TaTatacacaag	minus
					a = 52%	
		FOXC1	Forkhead box C1	2	cttgtGTA	plus

Continued

					t = 31%	
		FOXL1	Forkhead box L1	5	tgtgtATA	plus
					t = 48%	
		FOXL1	Forkhead box L1	5	tatacAcA	minus
					A = 96%	
		FOXO3	Forkhead box O3	5	tatAcACA	minus
					A = 92%	
		HLTF	Helicase-like transcription factor	2	ttcCtTgtgt	plus
					T = 100%	
		SOX10	SRY (sex determining region Y)-box 10	27	cctTgT	plus
					T = 100%	
		SOX10	SRY (sex determining region Y)-box 10	50	ttgTgT	plus
					t = 45%	
		SOX17	SRY (sex determining region Y)-box 17	4	ttccTTGTg	plus
					T = 100%	
		SRY	Sex determining region Y	1	gtagACAAt	minus
					A = 100%	
		TBP	TATA box binding protein	1	ctATAtAcacaagga	minus
					a = 22%	
rs1877474	**C**	HLTF	Helicase-like transcription factor	7	agcCtTttcg	plus
(**C**/T)					c = 24%	
		SPIB	Spi-B transcription factor (Spi-1/PU.1 related)	12	agcGaAA	minus
					G = 96%	
	T	ETS1	Protein C-ets-1	54	ttTgCt	plus
					T = 98%	
		HLTF	Helicase-like transcription factor	1	agcCtTttg	plus
					t = 14%	
		FOXA1	Forkhead box A1	1	TtTTTgCtctg	plus
					T = 91%	
		SOX10	SRY (sex determining region Y)-box 10	119	cttTtT	plus
					T = 95%	
		SPIB	Spi-B transcription factor (Spi-1/PU.1 related)	12	acaGcAA	minus
					A = 98%	
rs11119982	**C**	ARNT:AHR	aryl hydrocarbon receptor nuclear translocator	10	cGaGTG	minus
(C/T)			aryl hydrocarbon receptor		G = 96%	
		GATA2	GATA binding protein 2	23	cGATg	plus
					c = 25%	
		GATA3	GATA binding protein 3	4	cGATgg	plus

Continued

				c = 22%	
	HLTF	Helicase-like transcription factor	1	ggcCaTcgag	minus
				g = 37%	
	NKX3-2	Natural killer 3 homeobox 2	1	tcgAGTgtc	minus
				g = 13%	
	PAX2	Paired box gene 2	3	agaCActc	plus
				c = 35%	
	TFAP2A	Transcription factor AP-2 alpha	1	GCCatcgag	minus
		(activating enhancer binding protein 2 alpha)		G = 43%	
T	ARID3A	AT rich interactive domain 3A (BRIGHT-like)	30	ATcAAg	minus
				A = 100%	
	ARNT	aryl hydrocarbon receptor nuclear translocator	33	cAaGTG	minus
				A = 95%	
	ARNT	aryl hydrocarbon receptor nuclear translocator	33	cActTG	plus
				A = 100%	
	GATA2	GATA binding protein 2	136	tGATg	plus
				t = 17%	
	GATA3	GATA binding protein 3	342	tGATgg	plus
				t = 32%	
	MAX	MYC associated factor X	5	agaCACtTGa	plus
				T = 100%	
	MYB	v-*myb* myeloblastosis viral oncogene homolog	1	gaCacTTg	plus
				T = 100%	
	NKX2-5	Natural killer 2 homeobox 5	13	tcAAgtg	minus
				A = 100%	
	NKX3-2	Natural killer 3 homeobox 2	6	tcaAGTgtc	minus
				a = 8%	
	PAX2	Paired box gene 2	7	agaCactt	plus
				t = 26%	
	PAX2	Paired box gene 2	2	catCaagt	minus
				a = 84%	
	TFAP2A	Transcription factor AP-2 alpha	4	GCCatcaag	minus
		(activating enhancer binding protein 2 alpha)		a = 29%	
	USF1	upstream transcription factor 1	7	CAaGTGt	minus
				A = 100%	
	USF1	upstream transcription factor 1	13	CACtTGa	plus
				T = 93%	
	ZEB1	Zinc finger E-box binding homeobox 1	33	cACTTg	plus
				T = 98%	

activation transcription factor/cAMP responsive element-binding (CREB) protein family of transcription factors. This gene is one of seven ATFs that influence cellular physiologic processes by regulating the expression of downstream target genes, which are related to growth, survival, and other cellular activities. The three *ATF*3 rSNPs [rs3125289 (C/T), rs1877474 (C/T) and rs-11119982 (C/T)] in intron one has been found to be significantly associated with the risk of hypospadias [6]. The rs3125289 and rs1877474 AFT3-T alleles have a significantly greater occurrence in the case subjects compared to the corresponding control group where the rs11119982 *ATF*3-C allele has a significantly greater advantage [6].

The rs3125289 *ATF*3-C allele creates eight unique TFBS for the ARNT, ARNT:AHR, GABPA, MYC, MYCN, MZF1_1-4, SPIB and USF1 TFs which are involved in xenobiotic metabolism, cell cycle progression, apoptosis, cellular transformation, a variety of tumors, lymphoid-specific enhancer and cellular transcription, respectively (**Table 1**). The rs3125289 *ATF*3-T allele also creates eight unique TFBS for the FOXA1 & 2, FOXL1, FOXO3, HLTF, SOX10 & 17 and SRY TFs which are involved in embryonic development, apoptosis, chromatin structure, cell fate and male development, respectively (**Table 1** & **Supplement**). Three TFBS have been conserved between the two rs3125289 alleles which are for the ELF4, FOXC1 and TBP TFs that regulate epithelium-specific genes, resistance to oxidative stress and a core factor in DNA binding for TFIID, respectively (**Table 1**).

The rs1877474 *AFT*3-C allele creates no unique TFBS while the *AFT*3-T allele creates three unique TFBS which are for the ETS1, FOXA1 and SOX10 TFs which interact with TTRAP, UBE2I and Death associated proteins and regulates embryonic development, respectively (**Table 1**). Two TFBS have been conserved between the two alleles which are HLTF and SPIB TFs that are involved in altering chromatin structure and act as a lymphoid-specific enhancer, respectively (**Table 1**).

The rs11119982 *ATF*3-C allele creates one unique TFBS for the HLTF TF which is involved with altering chromatin structure. The *ATF*3-T allele creates five unique TFBS for the ARID3A, MAX, MYB, USF1 and ZEB1 TFs. These TFs are involved with the AT-rich interaction domain, transcription regulation, hematopoietic progenitor cells control, cellular transcription and repression, respectively (**Table 1**). Eight TFBS have been conserved between the two alleles which are ARNT, ARNT:AHR, GATA2 & 3, HKX2 & 3, PAX2 and TFAP2A TFs that are involved with xenobiotic metabolism, hematopoietic and endocrine cell lineages, endothelial cell boilogy, transcription repression, kidney cell differentiation and transcription activation, respectively (**Table 1**).

4. DISCUSSION

GWAS over the last decade have identified nearly 6500 disease or trait-predisposing SNPs where only 7% of these are located in protein-coding regions of the genome [23,24] and the remaining 93% are located within non-coding areas [25,26] 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, which are drawing more attention [7, 9,27]. 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 the 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 [27]. Examples of rSNPs associated with disease susceptibility are numerous and several reviews have been published [27-30].

In this report, the rs3125289 *ATF*3-C allele [C (+ strand) or G (−strand)] located in the ARNT and ARNT:AHR TFBS has a 100% and 96% occurrence, respectively, in humans; however, these BS occurs seven times in the gene and therefore a change in the TFBS created by the rSNP would probably not have any impact on the regulation of the gene (**Table 1** & **Supplement**). The *ATF*3-C allele is also located in the MYC and MYCN TFBS with a 90% occurrence and this BS occurs only once in the gene and therefore a nucleotide change created by the rSNP on these TFBS would probably have an impact on the regulation of the gene. The rs3125289 *ATF*3-T allele located in the FOXA2 and SRY TFBS has a 100% occurrence in humans and each BS occurs only once in the gene; therefore, the rSNP would probably have an impact on the regulation of the gene. Since the SRY TF located on the Y chromosome is a transcriptional regulator that controls a genetic switch in male development, this rSNP might be expected to have an impact on male etiology as has been shown to be the case with its risk of hypospadias [6].

The rs1877474 *ATF*3-T allele [A (−strand) or T (+ strand)] located in the FOXA1 TFBS has an occurrence of 91% in humans and this BS occurs only once in the gene and thereby might have an impact on embryonic development and tissue differentiation (**Table 1**). The other two unique TFBS (ETS1 and SOX10) occur multiple times in the gene and would not be expected to have an impact on gene regulation. The rs1877474 *ATF*3-C

allele does not have any unique TFBS and therefore would not be expected to have any impact on gene regulation.

The rs11119982 *ATF*3-C allele [C (+strand) or G (−strand)] located in the HLTF TFBS has an occurrence of 37% in humans (table). While this BS motif occurs only once in the gene and other HLTF BS motifs are found in intron one, it is doubtful that a change in the BS created by the rSNP would have an impaction on gene regulation. The rs11119982 *ATF*3-T allele located in the MYB TFBS has a 100% nucleotide occurrence for that BS motif in humans and occurs only once in the gene and thereby may have an impact on gene regulation and in the control of proliferation and differentiation of hematopoietic progenitor cells. Other rs11119982 *ATF*3-T allele unique TFBS affected by this rSNP occur multiple times in the gene and would probably have little effect on gene regulation. Since each of these rSNPs can alter TFBS in *ATF*3 intron one, it is easy to understand why the rSNPs have been found to independently affect the risk of hypospadias [6].

Human diseases or conditions that have been significantly associated with these rSNPs of the *ATF*3 gene are shown in the table along with rSNP allele-specific TFBS. 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. This change in the regulatory landscape can alter gene regulation which in turn can result in human disease, a change in condition or illness. In this report, three intron one rSNPs of the *ATF*3 gene have been described to illustrate that a change in rSNP alleles can provide different TFBS which in turn are also significantly associated with human etiology.

REFERENCES

[1] Chen, B.P., Liang, G., Whelan, J. and Hai, T. (1994) ATF3 and ATF3 delta Zip. Transcriptional repression versus activation by alternatively spliced isoforms. *The Journal of Biological Chemistry*, **269**, 15819-15826.

[2] Hai, T. and Hartman, M.G. (2001) The molecular biology and nomenclature of the activating transcription factor/cAMP responsive element binding family of transcription factors: Activating transcription factor proteins and homeostasis. *Gene*, **273**, 1-11. http://dx.doi.org/10.1016/S0378-1119(01)00551-0

[3] Thompson, M.R., Xu, D. and Williams, B.R. (2009) ATF3 transcription factor and its emerging roles in immunity and cancer. *Journal of Molecular Medicine*, **87**, 1053-1060. http://dx.doi.org/10.1007/s00109-009-0520-x

[4] Hai, T., Wolford, C.C. and Chang, Y.S. (2010) ATF3, a hub of the cellular adaptive-response network, in the pathogenesis of diseases: is modulation of inflammation a unifying component? *Gene Expression*, **15**, 1-11. http://dx.doi.org/10.3727/105221610X12819686555015

[5] Liu, B., Agras, K., Willingham, E., Vilela, M.L. and Baskin, L.S. (2006) Activating transcription factor 3 is estrogen-responsive in utero and upregulated during sexual differentiation. *Hormone Research*, **65**, 217-222. http://dx.doi.org/10.1159/000092402

[6] Beleza-Meireles, A., Tohonen, V., Soderhall, C., Schwentner, C., Radmayr, C., Kockum, I. and Nordenskjold, A. (2008) Activating transcription factor 3: A hormone responsive gene in the etiology of hypospadias. *European Journal of Endocrinology*, **158**, 729-739. http://dx.doi.org/10.1530/EJE-07-0793

[7] Knight, J.C. (2003) Functional implications of genetic variation in non-coding DNA for disease susceptibility and gene regulation. *Clinical Science*, **104**, 493-501. http://dx.doi.org/10.1042/CS20020304

[8] Knight, J.C. (2005) Regulatory polymorphisms underlying complex disease traits. *Journal of Molecular Medicine*, **83**, 97-109. http://dx.doi.org/10.1007/s00109-004-0603-7

[9] Wang, X., Tomso, D.J., Liu, X. and Bell, D.A. (2005) Single nucleotide polymorphism in transcriptional regulatory regions and expression of environmentally responsive genes. *Toxicology and Applied Pharmacology*, **207**, 84-90. http://dx.doi.org/10.1016/j.taap.2004.09.024

[10] Wang, X., Tomso, D.J., Chorley, B.N., Cho, H.Y., Cheung, V.G., Kleeberger, S.R. and Bell, D.A. (2007) Identification of polymorphic antioxidant response elements in the human genome. *Human Molecular Genetics*, **16**, 1188-1200. http://dx.doi.org/10.1093/hmg/ddm066

[11] Claessens, F., Verrijdt, G., Schoenmakers, E., Haelens, A., Peeters, B., Verhoeven, G. and Rombauts, W. (2001) Selective DNA binding by the androgen receptor as a mechanism for hormone-specific gene regulation. *The Journal of Steroid Biochemistry and Molecular Biology*, **76**, 23-30. http://dx.doi.org/10.1016/S0960-0760(00)00154-0

[12] Hsu, M.H., Savas, U., Griffin, K.J. and Johnson, E.F. (2007) Regulation of human cytochrome P450 4F2 expression by sterol regulatory element-binding protein and lovastatin. *The Journal of Biological Chemistry*, **282**, 5225-5236. http://dx.doi.org/10.1074/jbc.M608176200

[13] Takai, H., Araki, S., Mezawa, M., Kim, D.S., Li, X., Yang, L., Li, Z., Wang, Z., Nakayama, Y. and Ogata, Y. (2008) AP1 binding site is another target of FGF2 regulation of bone sialoprotein gene transcription. *Gene*, **410**, 97-104. http://dx.doi.org/10.1016/j.gene.2007.11.017

[14] Buroker, N.E., Huang, J.Y., Barboza, J., Ledee, D.R., Eastman, R.J., Jr., Reinecke, H., Ning, X.H., Bassuk, J.A. and Portman, M.A. (2012) The adaptor-related protein complex 2, alpha 2 subunit (AP2alpha2) gene is a peroxisome proliferator-activated receptor cardiac target gene. *The Protein Journal*, **31**, 75-83. http://dx.doi.org/10.1007/s10930-011-9379-0

[15] Huang, C.N., Huang, S.P., Pao, J.B., Hour, T.C., Chang, T.Y., Lan, Y.H., Lu, T.L., Lee, H.Z., Juang, S.H., Wu, P.P., Huang, C.Y., Hsieh, C.J. and Bao, B.Y. (2012) Genetic polymorphisms in oestrogen receptor-binding sites affect clinical outcomes in patients with prostate cancer receiving androgen-deprivation therapy. *Journal of Inter-*

nal Medicine, **271**, 499-509. http://dx.doi.org/10.1111/j.1365-2796.2011.02449.x

[16] Huang, C.N., Huang, S.P., Pao, J.B., Chang, T.Y., Lan, Y.H., Lu, T.L., Lee, H.Z., Juang, S.H., Wu, P.P., Pu, Y.S., Hsieh, C.J. and Bao, B.Y. (2012) Genetic polymorphisms in androgen receptor-binding sites predict survival in prostate cancer patients receiving androgen-deprivation therapy. *Annals of Oncology*, **23**, 707-713. http://dx.doi.org/10.1093/annonc/mdr264

[17] Yu, B., Lin, H., Yang, L., Chen, K., Luo, H., Liu, J., Gao, X., Xia, X. and Huang, Z. (2012) Genetic variation in the Nrf2 promoter associates with defective spermatogenesis in humans. *Journal of Molecular Medicine*.

[18] Wu, J., Richards, M.H., Huang, J., Al-Harthi, L., Xu, X., Lin, R., Xie, F., Gibson, A.W., Edberg, J.C. and Kimberly, R.P. (2011) Human FasL gene is a target of beta-catenin/T-cell factor pathway and complex FasL haplotypes alter promoter functions. *PLoS One*, **6**, e26143. http://dx.doi.org/10.1371/journal.pone.0026143

[19] Alam, M., Pravica, V., Fryer, A.A., Hawkins, C.P. and Hutchinson, I.V. (2005) Novel polymorphism in the promoter region of the human nerve growth-factor gene. *International Journal of Immunogenetics*, **32**, 379-382. http://dx.doi.org/10.1111/j.1744-313X.2005.00541.x

[20] Bryne, J.C., Valen, E., Tang, M.H., Marstrand, T., Winther, O., da Piedade, I., Krogh, A., Lenhard, B. and Sandelin, A. (2008) JASPAR, the open access database of transcription factor-binding profiles: New content and tools in the 2008 update. *Nucleic Acids Research*, **36D**, 102-106. http://dx.doi.org/10.1093/nar/gkm955

[21] Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W.W. and Lenhard, B. (2004) JASPAR: An open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Research*, **32D**, 91-94. http://dx.doi.org/10.1093/nar/gkh012

[22] Sandelin, A., Wasserman, W.W. and Lenhard, B. (2004) ConSite: web-based prediction of regulatory elements using cross-species comparison. *Nucleic Acids Research*, **32**, W249-252.

[23] Pennisi, E. (2011) The Biology of Genomes. Disease risk links to gene regulation. *Science*, **332**, 1031. http://dx.doi.org/10.1126/science.332.6033.1031

[24] Kumar, V., Wijmenga, C. and Withoff, S. (2012) From genome-wide association studies to disease mechanisms: Celiac disease as a model for autoimmune diseases. *Seminars in Immunopathology*, **34**, 567-580. http://dx.doi.org/10.1007/s00281-012-0312-1

[25] Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S. and Manolio, T.A. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences*, **106**, 9362-9367. http://dx.doi.org/10.1073/pnas.0903103106

[26] Kumar, V., Westra, H.J., Karjalainen, J., Zhernakova, D.V., Esko, T., Hrdlickova, B., Almeida, R., Zhernakova, A., Reinmaa, E., Vosa, U., Hofker, M.H., Fehrmann, R.S., Fu, J., Withoff, S., Metspalu, A., Franke, L. and Wijmenga, C. (2013) Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. *PLoS Genetics*, **9**, e1003201. http://dx.doi.org/10.1371/journal.pgen.1003201

[27] Chorley, B.N., Wang, X., Campbell, M.R., Pittman, G.S., Noureddine, M.A. and Bell, D.A. (2008) Discovery and verification of functional single nucleotide polymerphisms in regulatory genomic regions: current and developing technologies. *Mutation Research*, **659**, 147-157. http://dx.doi.org/10.1016/j.mrrev.2008.05.001

[28] Prokunina, L. and Alarcon-Riquelme, M.E. (2004) Regulatory SNPs in complex diseases: Their identification and functional validation. *Expert Reviews in Molecular Medicine*, **6**, 1-15. http://dx.doi.org/10.1017/S1462399404007690

[29] Buckland, P.R. (2006) The importance and identification of regulatory polymorphisms and their mechanisms of action. *Biochimica et Biophysica Acta*, **1762**, 17-28. http://dx.doi.org/10.1016/j.bbadis.2005.10.004

[30] Sadee, W., Wang, D., Papp, A.C., Pinsonneault, J.K., Smith, R.M., Moyer, R.A. and Johnson, A.D. (2011) Pharmacogenomics of the RNA world: Structural RNA polymorphisms in drug therapy. *Clinical Pharmacology & Therapeutics*, **89**, 355-365. http://dx.doi.org/10.1038/clpt.2010.314

Supplement. TF discription.

TFs	TF discription
ARID3A	This gene encodes a member of the ARID (AT-rich interaction domain) family of DNA binding proteins.
ARNT	Involved in the induction of several enzymes that participate in xenobiotic metabolism.
ARNT:AHR	The dimer alters transcription of target genes. Involved in the induction of several enzymes that participate in xenobiotic metabolism.
ELF5	The protein encoded by this gene is a member of an epithelium-specific subclass of the Etstranscritpion factor family. In addition to its role in regulating the later stages of terminal differentiation of keratinocytes, it appears to regulate a number of epithelium-specific genes found in tissues containing glandular epithelium such as salivary gland and prostate.
ETS1	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 proteins.
FOXA1	Transcription factor that is involved in embryonic development, establishment of tissue-specific gene expression and regulation of gene expression in differentiated tissues.
FOXA2	Transcription factor that is 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.
FOXL1	This gene encodes a member of the forkhead/winged helix-box (FOX) family of transcription factors.
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.
GABPA	This gene encodes one of three GA-binding protein transcription factor subunits which functions as a DNA-binding subunit.
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.
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.
MAX	The protein encoded by this gene is a member of the basic helix-loop-helix leucine zipper (bHLHZ) family of transcription factors. Transcription regulator.
MYB	This gene encodes a transcription factor that is a member of the MYB family of transcription factor genes. Transcriptional activator and plays an important role in the control of proliferation and differentiation of hematopoietic progenitor cells.
MYC	The protein encoded by this gene is a multifunctional, nuclear phosphoprotein that plays a role in cell cycle progression, apoptosis and cellular transformation.
MYCN	This gene is a member of the MYC family and encodes a protein with a basic helix-loop-helix (bHLH) domain. Amplification of this gene is associated with a variety of tumors, most notably neuroblastomas.
MZF1_1-4	Trancription regulator
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.
NKX3-2	This gene encodes a member of the NK family of homeobox-containing proteins. Transcriptional repressor that acts as a negative regulator of chondrocyte maturation.
PAX2	Probable transcription factor that may have a role in kidney cell differentiation.
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.
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.
SRY	Transcriptional regulator that controls a genetic switch in male development.

Continued

TFAP2A	The protein encoded by this gene is a transcription factor that binds the consensus sequence 5'-GCCNNNGGC-3' and activates the transcription of some genes while inhibiting the transcription of others.
TBP	General transcription factor that functions at the core of the DNA-binding multiprotein factor TFIID.
USF1	This gene encodes a member of the basic helix-loop-helix leucine zipper family, and can function as a cellular transcription factor.
ZEB1	This gene encodes a zinc finger transcription factor. Acts as a transcriptional repressor.

