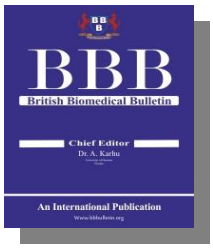
Journal Home Page www.bbbulletin.org

BRITISH BIOMEDICAL BULLETIN



Original

LIPA rSNPs (rs1412444 and rs2246833), Transcriptional Factor Binding Sites and Disease

Norman E. Buroker*

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

A R T I C L E I N F O

Received 05 June 2015
Received in revised form 18 June 2015
Accepted 21 June 2015

Keywords:
LIPA,
rSNPs,
TFBS,
TFs,
Disease.

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

A B S T R A C T

Background: The lysosomal acid lipase A (*LIPA*) gene transcribes the lysosomal acid lipase (LAL) which hydrolyzes cholesteryl esters and triglycerides in the cell lysosome thereby generating free cholesterol and fatty acids. Two *LIPA* single nucleotide polymorphisms (SNPs, rs1412444 and rs2246833) found to be associated with coronary artery disease (CAD) and myocardial infarction (MI) were analyzed with respect to alterations in transcriptional factor binding sites (TFBS) created by the SNP alleles.

Materials & Methods: The JASPAR CORE and ConSite databases were used to identify the TFBS. The Vector NTI Advance 11.5 computer program (Invitrogen, Life Technologies) was used to identify TFBS in the *LIPA* gene.

Results: These regulatory (r) SNPs in intron one (rs2246833) and intron two (rs141244) which are in linkage disequilibrium were found to alter TFBS resulting in potential changes in *LIPA* regulation. The rs2246833 common *LIPA*-C allele creates three unique punitive TFBS for the FOXC1, SP2 and ZNF143 transcriptional factors (TFs), while the minor *LIPA*-T allele creates one unique punitive TFBS for the MZF1 TF. The rs1412444 common *LIPA*-C allele creates five unique punitive TFBS for the ELF1, ETS1, GABPA, HOXA5 and SPI1 TFs, while the minor *LIPA*-T allele creates twelve unique punitive TFBS for the FOXA1, FOXL1, FOXO3, HNF1B, MEF2A, NFIC, NFKB1, PAX2, SOX6, SOX9, SRY and THAP1 TFs.

Conclusion: The changes created by *LIPA*rs1412444 and rs2246833 SNP alleles provide alternative TFBS for TFs that would affect *LIPA* regulation which in turn can be associated with human disease. Between the two *LIPA* rs141244 alleles there are a total of 43 potential TFBS which suggests that this is prominent binding site for TFs involved with *LIPA* gene regulation.

Introduction

The lysosomal acid lipase A (*LIPA*) gene on chromosome 10q23.31 encodes lysosomal acid lipase (LAL) which hydrolyzes cholesteryl esters and triglycerides in the cell lysosome to generate free cholesterol and fatty acids[1]. Alterations in the LAL enzyme activity could produce an accumulation of triglycerides and cholesterol esters in the cell which would result in foam cell formation, complement activation, an inflammation process and atherosclerotic plaque formation[2]. Un-esterified cholesterol is a distinguishing characteristic of atherosclerotic lesions[3]. Cholesteryl ester hydrolysis has been shown to be a critical step in the enzymatic modification of low-density lipoprotein (LDL) particles in the intima[4,5] where the particles create a risk for cardiovascular disease (CD).

Genome wide association studies (GWAS) have identified several genes associated with the risk of coronary artery disease (CAD) among different ethnic groups[6-11] where *LIPA* is one of these genes[12-14]. Further, the LIPA-T alleles for both the rs1412444 (intron 2) and rs2246833 (intron 1) single nucleotide polymorphisms (SNPs) have found to be associated with the risk of CAD[12-14]. Single nucleotide changes that affect gene expression by impacting gene regulatory sequences such as promoters, enhances, and silencers are known as regulatory SNPs (rSNPs)[15-18]. A rSNPs within a transcriptional factor binding site (TFBS) can change a transcriptional factor's (TF) ability to bind its TFBS[19-22] in which case the TF would be unable to effectively regulate its target gene[23-27]. This concept is examined for the two *LIPA* rSNPs which are separated by 2926bp and exon 2 but still happen to be in strong linkage disequilibrium (LD) ($r^2 = 0.937$)[12,13].

Materials and Methods

The JASPAR CORE database[28,29] and ConSite[30] were used to identify the TFBS in this study. JASPAR is a collection of transcription factor DNA-binding preferences used for scanning genomic sequences where ConSite is a web-based tool for finding cis-regulatory elements in genomic sequences. The TFBS and rSNP location within the binding sites have previously been discussed[31-33]. The Vector NTI Advance 11.5 computer program (Invitrogen, Life Technologies) was used to locate the TFBS in the *LIPA* gene (NCBI Ref Seq NM_000235.3) from 9.3kb upstream of the transcriptional start site to 859bp past the 3'UTR which represents a total of 16.9 kb. 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 & Supplement).

Results

The common rs2246833 SNP *LIPA*-C allele creates three unique potential TFBS for the FOXC1, SP2 and ZNF143TFs, which are involved with cell viability and resistance to oxidative stress in the eye, the nuclear matrix and transcriptional activation, respectively (Table, Figure 1, Supplement). The minor *LIPA*-T allele creates one unique potential TFBS for the MZF1 TF which functions as a transcription regulator (Table, supplement). There are also three conserved TBFS for the NFIC, RFX5 and THAP1 TFs which activate and regulate transcription (Table, supplement). The common rs1412444 SNP *LIPA*-C allele creates five unique potential TFBS for the ELF1, ETS1, GABPA, HOXA5 and SPI1 TFs which are

involved with transcription regulation, death associated proteins, nuclear control of mitochondrial function, gene expression and regulation of alternative splicing of target genes, respectively (Table, Figure 1, Supplement). The minor rs1412444 SNP *LIPA*-T allele creates twelve unique potential TFBS for the FOXA1, FOXL1, FOXO3, HNF1B, MEF2A, NFIC,NFKB1, PAX2, SOX6, SOX9, SRY and THAP1 TFs which are involved with development of multiple organ systems, gastrointestinal epithelium, apoptosis, the embryonic pancreaes, skeletal and cardiac muscle development, activating transcription, tumorigenesis and apoptosis, the kidney, maintenance of cardiac and skeletal muscle cells, male development and G1/S cell-cycle progression, respectively (Table, supplement). There are also thirteen conserved potential TBFS for the ELF5, ELK1 & 4, FEV, FLI1, FOXP2, GATA4, HNF4G, MEF2C, NFATC2, SPIB, SOX10 and TCF7L2 TFs which are involved with the epithelium, MAPK signaling, the serum response factor, repression, transcriptional activation, myocardial differentiation and function, steroid hormone receptor activation, cardiac morphogenesis and myogenesis, T cell transcription, embryonic development and cell fate, transcription suppression, respectively (Table, supplement). Between the two *LIPA* rs141244 alleles there are a total of 43 potential TFBS which suggests that this is prominent BS for TFs involved with *LIPA* gene regulation.

Discussion

The genome-wide association studies (GWAS) has over the past decade provided us with nearly 6,500 disease or trait-predisposing SNPs. Only seven percent of these SNPs are located in protein-coding regions of the genome[34,35] while the remaining 93% are located within non-coding regions[36,37] such as gene regulatory or intergenic areas of the genome. Much attention has been drawn to SNPs that occur in the putative regulatory of a gene where a single nucleotide change in the DNA sequence of a potential TF motif may affect the process of gene regulation[15,17,38]. A nucleotide change in a TFBS can have multiple consequences. Since a TF can usually recognize a number of different binding motifs in a gene, the SNP may not change the TFBS interaction with the TF and consequently not alter the process of gene expression. In other cases the nucleotide change may increase or decrease the TF's ability to bind DNA which would result in allele-specific gene expression. In some cases a nucleotide change may eliminate the natural binding motif or generate a new BS as a result the gene is no longer regulated by the original TF[31,33]. Therefore, functional rSNPs in TFBS may result in differences in gene expression, phenotypes and susceptibility to environmental exposure[38]. Examples of rSNPs associated with disease susceptibility are numerous and several reviews have been published[38-41].

The *LIPA* gene transcribes lysosomal acid lipase (LAL) which hydrolyzes cholesteryl esters and triglycerides in the lysosome of cells to generate free cholesterol and fatty acids. Mutations in this gene can result in Wolman disease and cholesteryl ester storage disease.

The *LIPA* SNPs (rs1412444 and rs2246833) which are in LD have been found to be significantly associated with coronary artery disease (CAD)[12,13]. The rs2246833 rSNP *LIPA*-C allele [C (- strand) or G (+ strand)] located in the unique potential FOXC1, SP2 and ZNF143 TFBS have a 13, 33 and 60% occurrence in humans, respectively (Table, Figure 1). Since these binding sites (BS) occurs only once in the gene except for the FOXC1

TFBS which occurs twice, the *LIPA*-C allele would probably have marginal impact on these specific TFs regulating the gene. Although the potential SP2 TFBS in the *LIPA* gene has a low occurrence in humans, the SP2 TF is associated with the nuclear matrix and can activate or repress gene expression (Supplement). The alternate rs2246833 rSNP *LIPA*-T allele [T (- strand) or A (+ strand)] located in the unique potential MZF1_1-4 TFBS has a 40% occurrence in humans and occurs 33 times in the gene therefore the rSNP would not be expected to have much impact on the TFs regulating the gene (Table). On the other hand, the rs1412444 rSNP *LIPA*-C allele [C (- strand) or G (+ strand)] located in the unique potential GABPA and HOXA5 TFBSs have a 92 and 81% occurrence in humans and occur only once in the gene; therefore, should have a greater impact on gene regulation by their respective TFs (Table, Figure 2). Especially since the GABPA TF is involved in the activation of cytochrome oxidase expression and nuclear control of mitochondrial function (Supplement). The rs1412444 rSNP *LIPA*-T allele [T (- strand) or A (+ strand)] located in the unique potential HNF1B, MEF2A and SOX9 TFBS have a 90, 68 and 93% occurrence in humans and only occur once in the gene; therefore, should have an impact on gene regulation by their respective TFs (Table). The MEF2A TF activates many muscle-specific growth factor-induced, and stress-induced genes (Supplement) while the HNF1B and SOX9 TFs are involved with nephron and embryonic development, respectively (Supplement). The unique potential GABPA TFBS provided by the rs1412444 rSNP common C allele and not present with the minor T allele is a BS for a TF that is involved with mitochondrial function. Consequently individuals carrying the rs1412444 rSNP *LIPA*-T allele maybe at risk for mitochondrial related diseases such as damage to cells of the brain, heart, liver, skeletal muscles, kidney and the endocrine and respiratory systems. In fact, the rs1412444 rSNP *LIPA*-T allele has been significantly associated with CAD[7,12-14] where the T allele frequency was found to change from 0.491 in a control group compared to 0.561 in the CAD group[13]. Of the thirteen conserved potential TFBS for both rs1412444 rSNP *LIPA* alleles (Results), the ELK4, FEV and FLI1 TFBS have no known nucleotide occurrence for the *LIPA* T allele in over three thousand observations (Table, JASPAR CORE database[28,29]) which would render the associated TFs an ineffective binding motif for individuals having the T allele. These three TFs are involved with serum response, transcriptional repression and activation, respectively, which should have an impact on human disease.

The rs2246833 rSNP *LIPA* T allele has also been found to be significantly associated with CAD[12,13] which is not surprising since the two rSNPs are strongly linked as indicated by the high LD value (r^2 = 0.937). Human diseases or conditions can be associated with the two rSNPs of the *LIPA* 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.

Conclusion

A change in the regulatory DNA landscape can alter gene regulation which in turn can result in human disease, a change in condition or illness. In this report, examples have been described to illustrate that the *LIPA* rs1412444 and rs2246833 rSNP alleles can provide alternative TFBS for TFs that would alter *LIPA* regulation which in turn can be associated with human disease.

Appendix

Supplemental material is available for this article.

Conflict of interest

The author declares no conflict of interest.

References

1. Brown, M.S. & Goldstein, J.L. A receptor-mediated pathway for cholesterol homeostasis. *Science.* 232, 34-47 (1986).
2. Zschenker, O., Illies, T. & Ameis, D. Over expression of lysosomal acid lipase and other proteins in atherosclerosis. *J Biochem.* 140, 23-38 (2006).
3. Kruth, H.S. Localization of unesterified cholesterol in human atherosclerotic lesions. Demonstration of filipin-positive, oil-red-O-negative particles. *Am J Pathol.* 114, 201-8 (1984).
4. Bhakdi, S. *et al*. On the pathogenesis of atherosclerosis: enzymatic transformation of human low density lipoprotein to an atherogenic moiety. *J Exp Med.* 182, 1959-71 (1995).
5. Bhakdi, S., Lackner, K.J., Han, S.R., Torzewski, M. & Husmann, M. Beyond cholesterol: the enigma of atherosclerosis revisited. *Thromb Haemost.* 91, 639-45 (2004).
6. Deloukas, P. *et al*. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet.* 45, 25-33.
7. Schunkert, H. *et al*. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet.* 43, 333-8.
8. O'Donnell, C.J. *et al*. Genome-wide association study for coronary artery calcification with follow-up in myocardial infarction. *Circulation.* 124, 2855-64.
9. Lee, J.Y. *et al*. A genome-wide association study of a coronary artery disease risk variant. *J Hum Genet.* 58, 120-6.
10. Lu, X. *et al*. Genome-wide association study in Han Chinese identifies four new susceptibility loci for coronary artery disease. *Nat Genet.* 44, 890-4.
11. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nat Genet.* 43, 339-44.
12. Wild, P.S. *et al*. A genome-wide association study identifies LIPA as a susceptibility gene for coronary artery disease. *Circ Cardiovasc Genet.* 4, 403-12.
13. Vargas-Alarcon, G. *et al*. Single nucleotide polymorphisms within LIPA (Lysosomal Acid Lipase A) gene are associated with susceptibility to premature coronary artery disease. a replication in the genetic of atherosclerotic disease (GEA) Mexican study. *PLoS One.* 8, e74703.
14. Wang, Y. *et al*. Genetic variants associated with myocardial infarction and the risk factors in Chinese population. *PLoS One.* 9, e86332.
15. Knight, J.C. Functional implications of genetic variation in non-coding DNA for disease susceptibility and gene regulation. *Clin Sci* (*Lond)* 104, 493-501 (2003).
16. Knight, J.C. Regulatory polymorphisms underlying complex disease traits. *J Mol Med (Berl)* 83, 97-109 (2005).
17. Wang, X., Tomso, D.J., Liu, X. & Bell, D.A. Single nucleotide polymorphism in transcriptional regulatory regions and expression of environmentally responsive genes. *Toxicol Appl Pharmacol.* 207, 84-90 (2005).
18. Wang, X. *et al*. Identification of polymorphic antioxidant response elements in the human genome. *Hum Mol Genet.* 16, 1188-200 (2007).
19. Claessens, F. *et al*. Selective DNA binding by the androgen receptor as a mechanism for hormone-specific gene regulation. *J Steroid Biochem Mol Biol.* 76, 23-30 (2001).
20. Hsu, M.H., Savas, U., Griffin, K.J. & Johnson, E.F. Regulation of human cytochrome P450 4F2 expression by sterol regulatory element-binding protein and lovastatin. *J Biol Chem.* 282, 5225-36 (2007).
21. Takai, H. *et al*. AP1 binding site is another target of FGF2 regulation of bone sialoprotein gene transcription. *Gene.* 410, 97-104 (2008).

22. Buroker, N.E. *et al*. The adaptor-related protein complex 2, alpha 2 subunit (AP2alpha2) gene is a peroxisome proliferator-activated receptor cardiac target gene. *Protein J.* 31, 75-83 (2012).
23. Huang, C.N. *et al*. Genetic polymorphisms in oestrogen receptor-binding sites affect clinical outcomes in patients with prostate cancer receiving androgen-deprivation therapy. *J Intern Med.* 271, 499-509 (2012).
24. Huang, C.N. *et al*. Genetic polymorphisms in androgen receptor-binding sites predict survival in prostate cancer patients receiving androgen-deprivation therapy. *Ann Oncol.* 23, 707-13 (2012).
25. Yu, B. *et al*. Genetic variation in the Nrf2 promoter associates with defective spermatogenesis in humans. *J Mol Med (Berl)* (2012).
26. Wu, J. *et al*. Human FasL gene is a target of beta-catenin/T-cell factor pathway and complex FasL haplotypes alter promoter functions. *PLoS One.* 6, e26143 (2011).
27. Alam, M., Pravica, V., Fryer, A.A., Hawkins, C.P. & Hutchinson, I.V. Novel polymorphism in the promoter region of the human nerve growth-factor gene. *Int J Immunogenet.* 32, 379-82 (2005).
28. Bryne, J.C. *et al*. JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic Acids Res.* 36, D102-6 (2008).
29. Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W.W. & Lenhard, B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res.* 32, D91-4 (2004).
30. Sandelin, A., Wasserman, W.W. & Lenhard, B. ConSite: web-based prediction of regulatory elements using cross-species comparison. *Nucleic Acids Res.* 32, W249-52 (2004).
31. Buroker, N.E. ADRBK1 (GRK2) rSNPs, Transcriptional Factor Binding Sites and Cardiovascular Disease in the Black Population. *Journal of Cardiovascular Disease.* 2(2014).
32. Buroker, N.E. *et al*. AKT3, ANGPTL4, eNOS3, and VEGFA associations with high altitude sickness in Han and Tibetan Chinese at the Qinghai-Tibetan Plateau. *Int J Hematol.* 96, 200-13 (2012).
33. Buroker, N.E. AKT3 rSNPs, Transcritional Factor Binding Sites and Human Disease. *Open Journal of Blood Diseases.* 3, 116-129 (2013).
34. Pennisi, E. The Biology of Genomes. Disease risk links to gene regulation. *Science.* 332, 1031 (2011).
35. Kumar, V., Wijmenga, C. & Withoff, S. From genome-wide association studies to disease mechanisms: celiac disease as a model for autoimmune diseases. *Semin Immunopathol.* 34, 567-80 (2012).
36. Hindorff, L.A. *et al*. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci. U S A* 106, 9362-7 (2009).
37. Kumar, V. *et al*. Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. *PLoS genetics.* 9, e1003201 (2013).
38. Chorley, B.N. *et al*. Discovery and verification of functional single nucleotide polymorphisms in regulatory genomic regions: current and developing technologies. *Mutat Res.* 659, 147-57 (2008).
39. Prokunina, L. & Alarcon-Riquelme, M.E. Regulatory SNPs in complex diseases: their identification and functional validation. *Expert Rev Mol Med.* 6, 1-15 (2004).
40. Buckland, P.R. The importance and identification of regulatory polymorphisms and their mechanisms of action. *Biochimica et biophysica acta.* 1762, 17-28 (2006).
41. Sadee, W. *et al*. Pharmacogenomics of the RNA world: structural RNA polymorphisms in drug therapy. *Clin Pharmacol Ther.* 89, 355-65.

Table. The LIPA rSNPs 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
rs2246833 (C/T) Intron 1	**C**	FOXC1	Forkhead box C1	2	gcccgGTA g=13%	plus
		NFIC	Nuclear factor 1 C-type	8	cgGGca c=40%	minus
		RFX5	Regulatory factor X, 5	1	cctaccgGgcACaaa c=22%	minus
		SP2	Specificity Protein 2	1	atcCtgCCtaccggg c=33%	minus
		THAP1	THAP domain containing, apoptosis associated protein 1	1	gtgCCcggt g=39%	plus
		ZNF143	Zinc finger protein 143	1	ttdtttCCAtccTgCctacc c=60%	minus
	T	MZF1_1-4	Myeloid zinc finger 1	33	tGGGcA t=40%	minus
		NFIC	Nuclear factor 1 C-type	33	tgGGca t=40%	minus
		RFX5	Regulatory factor X, 5	1	cctactgGgcACaaa t=78%	minus
		THAP1	THAP domain containing, apoptosis associated protein 1	1	gtgCCcagt a=24%	plus
rs1412444 (C/T) Intron 2	**C**	ELF1	E74-like factor 1 (ets domain transcription factor)	1	aacaccGGAAata c=34%	minus
		ELF5	E74-like factor 5 (ets domain transcription factor)	1	tattTCCgg g=25%	plus
		ELK1	ELK1, member of ETS oncogene family	1	acaccgGAaa c=86%	minus
		ELK4	ELK4, ETS-domain protein	1	ttatTTCCggt g=85%	plus
		ETS1	Protein C-ets-1	1	gttattTCCggtgtt g=8%	plus

	FEV	ETS oncogene family	1	ccGGAAat c=23%	minus
	FLI1	Fli-1 proto-oncogene, ETS transcription factor	1	accGGAAataa c=26%	minus
	FOXP2	Forkhead box P2	1	cttTAAACAcc c=16%	minus
	GABPA	GA binding protein transcription factor	1	cCGGAAAtaac G=92%	minus
	GATA4	GATA binding protein 4	1	tgTTATttccg g=27%	plus
	HNF4G	hepatocyte nuclear factor 4	1	ggtgtttAaAGgcaa g=21%	plus
	HOXA5	Homeobox A5	1	cggaAATa c=81%	minus
	MEF2C	Myocyte enhancer factor 2C	1	acaccggAAATAaca c=45%	minus
	NFATC2	Nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 2	2	atTTCCg g=4%	plus
	SPIB	Spi-B transcription factor (Spi-1/PU.1 related)	1	accGGAA c=33%	minus
	SPI1	Spleen focus forming virus (SFFV) proviral integration oncogene	1	aaacaccGGAAtaa c=5%	minus
	SOX10	SRY (sex determining region Y)-box 10	3	cggTgT g=9%	plus
	TCF7L2	Transcription factor 7-like 2 (T-cell specific, HMG-box)	1	ggtgtTtAAAGgca g=27%	plus
T	ELF5	E74-like factor 5 (ets domain transcription factor)	4	tattTCCag a=36%	plus
	ELK1	ELK1, member of ETS oncogene family	2	acactgGAaa t=4%	minus
	ELK4	ELK4, ETS-domain protein	1	ttatTTCCagt a=0%	plus
	FEV	ETS oncogene family	5	ctGGAAat t=0%	minus
	FLI1	Fli-1 proto-oncogene, ETS transcription factor	1	actGGAAataa t=0%	minus
	FOXA1	Forkhead box A1	1	gtgtTaTTtaCagtg a=42%	plus
	FOXL1	Forkhead box L1	6	tggaaATA t=48%	minus

	FOXO3	Forkhead box O3	6	tggAAAtA t=62%	minus
	FOXP2	Forkhead box P2	1	cttTAAACAct t=16%	minus
	GATA4	GATA binding protein 4	1	tgTTATttcca a=14%	plus
	HNF1B	HNF1 homeobox B	1	ccAgTgTTTAAa A=90%	plus
	HNF4	Hepatocyte nuclear factor 4	1	agtgtttAaAGgcaa a=37%	plus
	MEF2A	Myocyte enhancer factor 2	1	cactggAAATAacac t=68%	minus
	MEF2C	Myocyte enhancer factor 2C	1	acactggAAATAaca t=15%	minus
	NFKB1	Nuclear factor of kappa light polypeptide gene enhancer in B-cells 1	6	tGttaTTtCCa a=31%	plus
	NFIC	Nuclear factor I/C (CCAAT-binding transcription factor)	39	cTGGaa T=94%	minus
	NFATC2	Nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 2	19	atTTCCa a=69%	plus
	PAX2	Paired box gene 2	2	aaaCactg t=32%	plus
	SOX6	SRY (sex determining region Y)-box 6	1	cCagTGTtta a=65%	plus
	SOX9	SRY (sex determining region Y)-box 9	1	ccAgtGttt A=93%	plus
	SOX10	SRY (sex determining region Y)-box 10	31	cagTgT a=36%	plus
	SPIB	Spi-B transcription factor (Spi-1/PU.1 related)	10	actGGAA t=4%	minus
	SRY	Sex determining region Y	1	ttaaACAct t=68%	minus
	TCF7L2	Transcription factor 7-like 2 (T-cell specific, HMG-box)	1	agtgtTtAAAGgca a=36%	plus
	THAP1	THAP domain containing, apoptosis associated protein 1	3	tttCCagtg a=29%	plus

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

TF	Protein name	TF description/function
ELF1	E74-like factor 1	This gene encodes an E26 transformation-specific related transcription factor. The encoded protein is primarily expressed in lymphoid cells and acts as both an enhancer and a repressor to regulate transcription of various genes.
ELF5	E74-like factor 5	A member of an epithelium-specific subclass of the Ets transcritpion factor family.
ELK1	ELK1, member of ETS oncogene family	This gene is a member of the Ets family of transcription factors and of the ternary complex factor (TCF) subfamily. The protein encoded by this gene is a nuclear target for the ras-raf-MAPK signaling cascade.
ELK4	ELK4, ETS-domain protein (SRF accessory protein 1)	This gene is a member of the Ets family of transcription factors and of the ternary complex factor (TCF) subfamily. Proteins of the TCF subfamily form a ternary complex by binding to the the serum response factor and the serum reponse element in the promoter of the c-fos proto-oncogene.
ETS1	Protein C-ets-1	The protein encoded by this gene belongs to the ETS family of transcription factors and has been shown to interact with TTRAP, UBE2I and Death associated protein.
FEV	ETS oncogene family	It functions as a transcriptional repressor.
FLI1	Fli-1 proto-oncogene, ETS transcription factor	Sequence-specific transcriptional activator. Recognizes the DNA sequence 5'-C[CA]GGAAGT-3'
FOXA1	Forkhead box A1	Transcription factor that is involved in embryonic development, establishment of tissue-specific gene expression and regulation of gene expression in differentiated tissues. Is thought to act as a 'pioneer' factor opening the compacted chromatin for other proteins through interactions with nucleosomal core histones and thereby replacing linker histones at target enhancer and/or promoter sites. Involved in the development of multiple endoderm-derived organ systems such as liver, pancreas, lung and prostate. Modulates the transcriptional activity of nuclear hormone receptors.
FOXA2	Forkhead box A2	Transcription factor that is involved in embryonic development, establishment of tissue-specific gene expression and regulation of gene expression in differentiated tissues.
FOXC1	Forkhead box C1	An important regulator of cell viability and resistance to oxidative stress in the eye.
FOXL1	Forkhead box L1	Transcription factor required for proper proliferation and differentiation in the gastrointestinal epithelium. Target gene of the hedgehog (Hh) signaling pathway.
FOXO3	Forkhead box O3	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.

FOXP2	Forkhead box P2	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.
GABPA	GA-binding protein alpha chain	One of three GA-binding protein transcription factor subunits which functions as a DNA-binding subunit which shares identity with a subunit encoding the nuclear respiratory factor 2 gene and is likely involved in activation of cytochrome oxidase expression and nuclear control of mitochondrial function.
GATA4	GATA binding protein 4	This gene encodes a member of the GATA family of zinc-finger transcription factors. Members of this family recognize the GATA motif which is present in the promoters of many genes. This protein is thought to regulate genes involved in embryogenesis and in myocardial differentiation and function. Mutations in this gene have been associated with cardiac septal defects.
HNF1B	HNF1 homeobox B	This gene encodes a member of the homeodomain-containing superfamily of transcription factors and has been shown to function in nephron development, and regulates development of the embryonic pancreas.
HNF4G	Hepatocyte nuclear factor 4, gamma	Steroid hormone receptor activity and sequence-specific DNA binding transcription factor activity. An important paralog of this gene is RXRA.
HOXA5	Homeobox protein Hox-A5	DNA-binding transcription factor which may regulate gene expression, morphogenesis, and differentiation.
MEF2A	Myocyte enhancer factor 2A	The protein encoded by this gene is a DNA-binding transcription factor that activates many muscle-specific, growth factor-induced, and stress-induced genes. Mediates cellular functions not only in skeletal and cardiac muscle development, but also in neuronal differentiation and survival.
MEF2C	Myocyte enhancer factor 2C	Transcription activator which binds specifically to the MEF2 element present in the regulatory regions of many muscle-specific genes. Controls cardiac morphogenesis and myogenesis, and is also involved in vascular development.
MZF1_1-4	Myeloid zinc finger 1	Binds to target promoter DNA and functions as trancription regulator. May be one regulator of transcriptional events during hemopoietic development. Isoforms of this protein have been shown to exist at protein level.
NFATC2	Nuclear factor of activated T-cells, cytoplasmic 2	This protein is present in the cytosol and only translocates to the nucleus upon T cell receptor (TCR) stimulation, where it becomes a member of the nuclear factors of activated T cells transcription complex.
NFIC	Nuclear factor 1 C-type	Recognizes and binds the palindromic sequence 5'-TTGGCNNNNNGCCAA-3' present in viral and cellular promoters and in the origin of replication of adenovirus type 2. These proteins are individually capable of activating transcription and replication.
NFKB1	Nuclear factor of kappa light polypeptide gene enhancer in B-cells	NF-kappa-B is a pleiotropic transcription factor present in almost all cell types and is the endpoint of a series of signal transduction

	1	events that are initiated by a vast array of stimuli related to many biological processes such as inflammation, immunity, differentiation, cell growth, tumorigenesis and apoptosis.
PAX2	Paired box gene 2	Probable transcription factor that may have a role in kidney cell differentiation.
RFX5	Regulatory factor X, 5	Activates transcription from class II MHC promoters. Recognizes X-boxes. Mediates cooperative binding between RFX and NF-Y. RFX binds the X1 box of MHC-II promoters
SOX6	SRY (sex determining region Y)-box 6	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. The encoded protein interacts with other family members to cooperatively activate gene expression.
SOX9	SRY (sex determining region Y)-box 9	The protein encoded by this gene recognizes the sequence CCTTGAG along with other members of the HMG-box class DNA-binding proteins.
SOX10	SRY (sex determining region Y)-box 10	This gene encodes a member of the SOX (SRY-related HMG-box) family of transcription factors involved in the regulation of embryonic development and in the determination of the cell fate.
SP2	Specificity Protein 2	This gene encodes a member of the Sp subfamily of Sp/XKLF transcription factors. Sp family proteins are sequence-specific DNA-binding proteins characterized by an amino-terminal trans-activation domain and three carboxy-terminal zinc finger motifs. This protein contains the least conserved DNA-binding domain within the Sp subfamily of proteins, and its DNA sequence specificity differs from the other Sp proteins. It localizes primarily within subnuclear foci associated with the nuclear matrix, and can activate or in some cases repress expression from different promoters.
SPIB	Transcription factor Spi-B	SPI1 and SPIB are members of a subfamily of ETS transcription factors. ETS proteins share a conserved ETS domain that mediates specific DNA binding. SPIB and SPI1 bind to a purine-rich sequence, the PU box (5-prime-GAGGAA-3-).
SPI1	Spleen focus forming virus (SFFV) proviral integration oncogene	This gene encodes an ETS-domain transcription factor that activates gene expression during myeloid and B-lymphoid cell development. The nuclear protein binds to a purine-rich sequence known as the PU-box found near the promoters of target genes, and regulates their expression in coordination with other transcription factors and cofactors. The protein can also regulate alternative splicing of target genes.
SRY	Sex determining region Y	Transcriptional regulator that controls a genetic switch in male development.
TCFCP211	Transcription factor CP2-like 1	Transcriptional suppressor. May suppress UBP1-mediated transcriptional activation. Modulates the placental expression of CYP11A1.
THAP1	THAP domain containing, apoptosis associated protein 1	DNA-binding transcription regulator that regulates endothelial cell proliferation and G1/S cell-cycle progression.

ZNF143	Zinc finger protein 143	Transcriptional activator. Activates the gene for selenocysteine tRNA (tRNAsec). Binds to the SPH motif of small nuclear RNA (snRNA) gene promoters. Participates in efficient U6 RNA polymerase III transcription via its interaction with CHD8

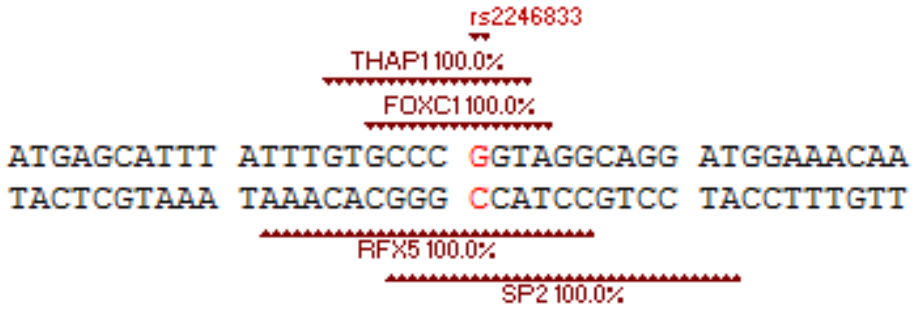


Figure 1. Double stranded DNA from the *LIPA* gene showing the potential TFBS for four different TFs which can bind their respective DNA sequence either above (+) or below (-) the duplex (cf. Table). The rs2246833 rSNP common *LIPA*-C allele is found in each of these TFBS. As shown, this rSNP is located in intron one of the *LIPA* gene. Also included with the potential TFBS is their % sequence homology to the duplex

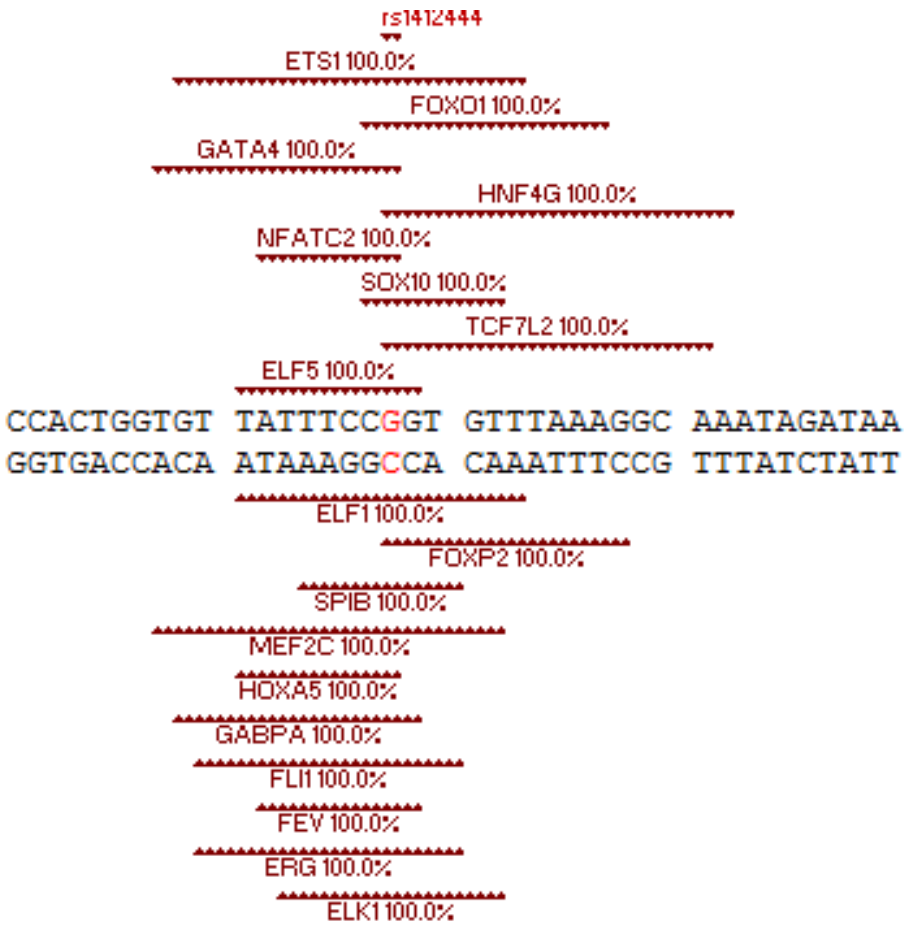


Figure 2. Double stranded DNA from the *LIPA* gene showing the potential TFBS for eighteen different TFs which can bind their respective DNA sequence either above (+) or below (-) the duplex (cf. Table). The rs1412444 common *LIPA*-C allele is found in each of these TFBS. As shown, this rSNP is located in intron two of the *LIPA* gene. Also included with the potential TFBS is their % sequence homology to the duplex