Frontiers in Bioscience Scholar

Front. Biosci. (Schol Ed) **2026**; 18(4): 51759
https://doi.org/10.31083/FBS51759

Original Research

Significant and Potentially Clinically Significant Genetic Variants Altering SREBP-Binding Sites Upstream of Human Genes From the Cholesterol Biosynthesis Regulatory Network

Elena Ignatieva[1,*], Angelina Mikhailova[2]

[1]Institute of Cytology and Genetics, SB RAS, 630090 Novosibirsk, Russia
[2]Department of Natural Sciences, Novosibirsk State University, 630090 Novosibirsk, Russia
*Correspondence: eignat@bionet.nsc.ru (Elena Ignatieva)

Academic Editor: Gustavo Caetano-Anollés
Submitted: 11 March 2026 Revised: 22 July 2026 Accepted: 27 July 2026 Published: 17 November 2026

Abstract

Background: Transcription factors from the sterol regulatory element-binding protein (SREBP) subfamily regulate the transcriptional activity of target genes in response to intracellular cholesterol levels. Genetic variants that colocalize with SREBP-binding sites may alter this regulation. Thus, this study aimed to assess the potential clinical significance of genetic variants identified in SREBP sites within the regulatory regions of human genes encoding enzymes of cholesterol biosynthesis. **Methods:** A collection of functionally active SREBP-binding sites was manually compiled based on data from research articles. The UCSC Table Browser data retrieval tool was used to extract data on genetic variants colocalized with SREBP sites from dbSNP155 and to obtain PhyloP score values from the UCSC Genome Browser database that characterized the evolutionary conservation of the nucleotides affected by genetic variants. The similarity of SREBP sites to the SREBP1- and SREBP2-binding motif models was estimated using MoloTool. **Results:** This study identified 171 single-nucleotide variants (SNVs) and 5 short deletions that were colocalized with 22 SREBP-binding sites. One of the five deletions (rs1563290033 in the *FDFT1* promoter) was a clinically significant variant previously identified in a patient with squalene synthase deficiency, a rare pathology. The *FDFT1* promoter containing this 16-nucleotide deletion activated *FDFT1* transcription less effectively than the wild-type promoter. To our knowledge, the colocalization of this deletion with the SREBP site has not been reported previously. Thus, to our knowledge, this study is the first to report that the rs1563290033 16-bp deletion leads to the loss of the SREBP 10-mer motif and the adjacent four and two flanking nucleotides. Five SNVs located in SREBP sites within the *SQLE*, *FDFT1*, and *SREBF2* promoter regions affected nucleotides with the highest level of evolutionary conservation (PhyloP score >4.6). According to analysis performed using MoloTool, three of these five SNVs (rs1680036166, rs1357951184, and rs1602252939) and four 1-bp deletions (rs1585844469, rs1029411034, rs1207705310, and rs1264806290) decreased the similarity of the SREBP sites to both SREBP-binding motif models. **Conclusions:** The rs1563290033 16-bp deletion associated with squalene synthase deficiency leads to the loss of the SREBP site from the *FDFT1* promoter, which explains the mechanism reducing the activity of the *FDFT1* promoter carrying this deletion. Notably, the three SNVs and four 1-bp deletions described above are promising candidates for further experimental testing of their effect on the functional activity of SREBP sites.

Keywords: cholesterol; lipid metabolism; gene regulatory networks; sterol regulatory element binding proteins; promoter regions; transcription factors; polymorphism; genetic; pharmacogenomic variants; genetic predisposition to disease

1. Introduction

In humans, cholesterol is an essential structural component of cell membranes and blood plasma lipoproteins, a precursor of a number of steroid hormones, bile acids and vitamin D [1]. Cholesterol can be absorbed from food [2] and synthesized *de novo* in a number of tissues and organs [3], including the liver, testes, spleen, adrenal gland, white and brown fat, kidney, heart, brain, muscles, skin, preputial tissue [4], astrocytes, and developing neurons [5]. The cholesterol biosynthesis pathway includes 21 stages and involves enzymes encoded by 24 genes [6,7]. The initial compounds for cholesterol biosynthesis are acetyl-CoA and acetoacetyl-CoA [6]. Cholesterol biosynthesis begins with the mevalonate pathway, which produces isoprenoids (isopentenyl pyrophosphate and dimethylallyl pyrophosphate): the intermediate products of the mevalonate pathway are mevalonic acid, mevalonate, and mevalonate diphosphate [7]. The next six reactions are carried out with the participation of five enzymes (FDPS, GGPS1, FDFT1, SQLE, LSS), and at this stage lanosterol is synthesized. Further, cholesterol synthesis can follow two alternative pathways: the Bloch pathway and the Kandutsch-Russell pathway, each of which includes eight stages [6,7]. Intermediate products of the cholesterol biosynthetic pathway also perform important functions in animal cells. In particular, geranylgeranyl pyrophosphate and farnesyl pyrophosphate are substrates in prenylation reactions. Many proteins undergo this posttranslational modification (for example, small GTP-binding proteins), which is important for the proper localization and activation of these proteins [8].

Copyright: © 2026 The Author(s). Published by IMR Press.
This is an open access article under the CC BY 4.0 license.

Publisher's Note: IMR Press stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

A vital rate-limiting enzyme within the mevalonate pathway is 3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMGCR). It converts 3-hydroxy-3-methyl-glutaryl coenzyme A (HMG-CoA) to mevalonate [9]. HMGCR is a target for statins, important cholesterol-lowering drugs used to treat cardiovascular disease, and therefore extensive efforts have been made to study HMGCR regulation and genetic variation. In mice, liver-specific *HMGCR* knockout causes hepatic steatosis and death [10]. Genetic variants of this gene are associated with a number of pathologies. Among them are three skeletal muscle diseases: limb-girdle muscular dystrophy, statin-associated myopathy, and statin-associated autoimmune myopathy (autoimmune anti-HMGCR myopathy) [9,11,12]. Genetic variants in the *HMGCR* gene have also been identified as affecting total cholesterol and low-density lipoprotein (LDL) cholesterol [13] and are associated with other multiple lipid/lipoprotein traits, especially with triglyceride, small LDL, and high-density lipoprotein (HDL) cholesterol [14].

Previously, we built a gene network regulating cholesterol biosynthesis [15,16]. The central players in this network are transcription factors from the sterol regulatory element-binding protein (SREBP) subfamily, whose activities are inversely regulated by intracellular cholesterol [17]. In humans, this subfamily includes SREBP1a, SREBP1c, and SREBP2, encoded by the *SREBF1* and *SREBF2* genes [18,19].

There are two feedback loops in the gene network regulating cholesterol biosynthesis (Fig. 1). First, the *SREBF2* gene encodes transcription factor SREBP2, which activates the transcription of its own gene [20]. The second feedback loop involves cholesterol, SREBP cleavage-activating protein (SCAP), proteins encoded by insulin induced genes (INSIG1 or INSIG2), SREBPs, their target genes, and cholesterol biosynthesis enzymes encoded by them. This feedback loop functions as follows. SREBPs are synthesized in a form of inactive precursors (preSREBPs, not shown in Fig. 1) associated with SCAP on the membrane of the endoplasmic reticulum [18,21]. At low cholesterol levels, INSIG proteins undergo ubiquitination and rapid degradation (not shown in Fig. 1), and the SCAP/preSREBPs complex is transported to the Golgi apparatus (not shown in Fig. 1), where the site-1 and site-2 proteases (S1P and S2P) (not shown in Fig. 1) cleave preSREBP [17]. As a result, the N-terminal domains of preSREBPs, which are active transcription factors, are released [17]. The SREBPs formed in this way are able to bind to their binding sites in the regulatory regions of target genes encoding cholesterol biosynthesis enzymes and activate their transcription [15]. Thus, the rate of cholesterol biosynthesis increases. At high cholesterol levels, the formation of the active form of SREBPs is inhibited, since cholesterol induces binding (not shown in Fig. 1) between SCAP and membrane proteins INSIGs. Due to INSIGs, the SCAP/preSREBPs complex is retained on the membrane (not shown in Fig. 1), thus preventing the formation of active SREBPs [17].

SREBPs belong to the basic-helix-loop-helix / leucine zipper class of transcription factors. Initially, the SREBP-binding sites were defined as ten-nucleotide DNA regions similar to the TCACNCCAC sequence motif [22,23]. It has been shown that SREBP1 has dual DNA sequence specificity, binding to both the classical sterol regulatory element (SRE, ATCACCCCAC) and the E-box motif (ATCACGTGA) [23]. Later, other models of SREBP-binding sites were proposed, which included more than ten nucleotides. For example, in the HOCOMOCO database [24], these are RTCASGTGAYbb (for SREBP1) and dRWSGCGTGAKn (for SREBP2).

Experiments on mouse lines (transgenic for SREBP-1a, transgenic for SREBP-2, and knockout for *Scap*) have shown that the expression of almost all the genes encoding enzymes of the cholesterol biosynthesis pathway is regulated by SREBPs [25,26]. When studying the regulatory regions of human genes, functionally active SREBP-binding sites have been found in the regulatory regions of a considerable part of genes from the gene network regulating cholesterol biosynthesis [6,16,27].

It is known that disorders of cholesterol biosynthesis are accompanied by severe pathologies (Smith-Lemli-Opitz Syndrome, desmosterolosis, lathosterolosis, Conradi–Hünermann syndrome, child syndrome, Greenberg skeletal dysplasia, X-linked chondrodysplasia punctata, mevalonic aciduria, hyperimmunoglobulinemia D syndrome, porokeratosis, etc.) [28,29,30,31]. Genetic variants in the genes encoding cholesterol biosynthesis enzymes associated with these diseases have been identified [13,32,33]. Most of these variants are in the coding parts of genes, and they lead to impairments of protein structure and decreased activity [34,35,36]. However, a decrease in protein activity may also be caused by reduced gene expression owing to alterations in regulatory elements (methylation and transcription factor binding sites, topologically associating domain (TAD) boundaries, etc.) in noncoding regions of the genome [37,38]. Such alterations may occur due to nucleotide substitutions, deletions and insertions.

The goals of this study are (1) identification of short genetic variants (single-nucleotide variants (SNVs), deletions) colocalized with experimentally studied SREBP-binding sites in regulatory regions of genes from the gene network controlling cholesterol biosynthesis and (2) evaluation of potential clinical significance of these genetic variants, taking into account the evolutionary conservation of affected nucleotides and the effects of variants on the similarity of the SREBP sites to SREBP1- and SREBP2-binding motif models.

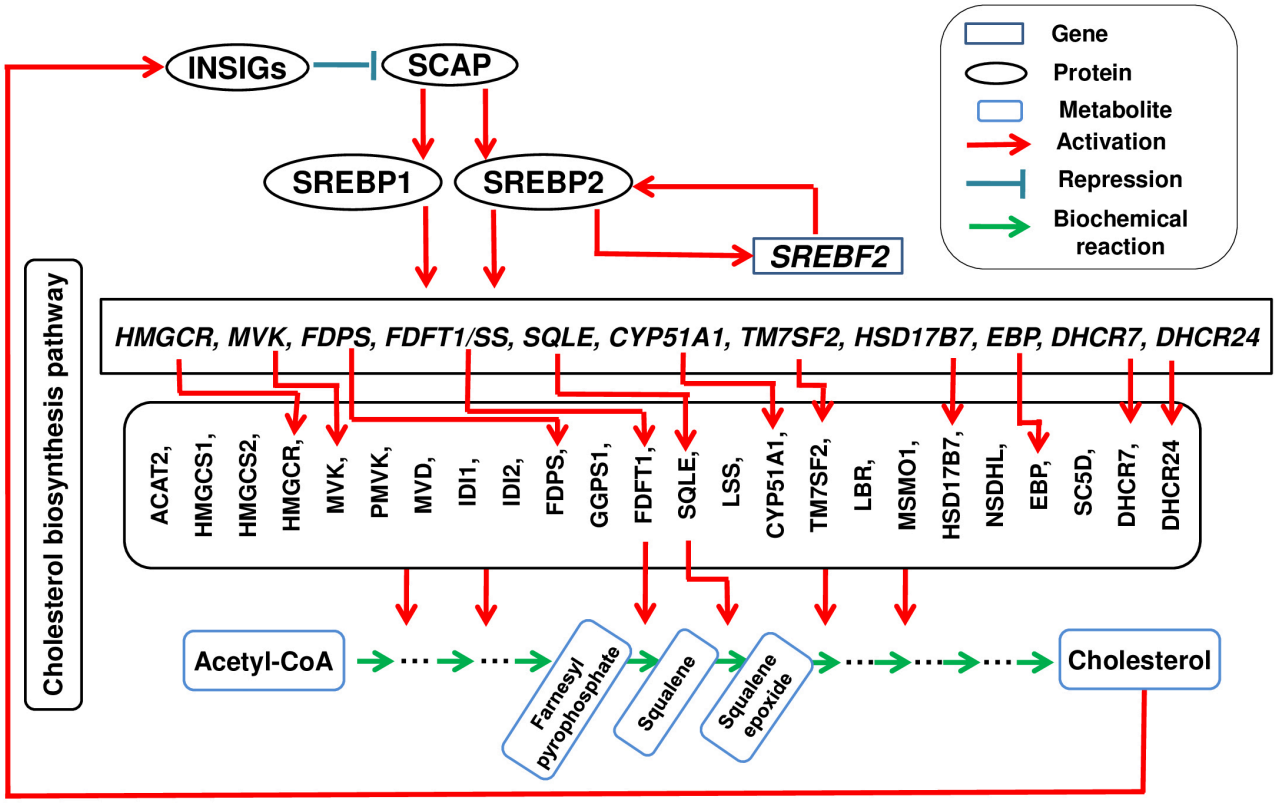


Fig. 1. Schematic representation of regulatory feedback loops operating in the gene network that controls cholesterol biosynthesis. The first feedback loop includes the *SREBF2* gene, encoding SREBP2, which activates transcription of its own gene. The second feedback loop includes cholesterol changing the INSIG and SCAP conformation. It thereby prevents SCAP from activating SREBP1 and SREBP2, prevents subsequent activation of SREBP target genes encoding enzymes involved in cholesterol biosynthesis and the activation of enzymes themselves (for a detailed description of these two feedback loops, see the text). Only genes with functionally active SREBP-binding sites in their regulatory regions (Table 1) are shown in this figure. The enzymes of cholesterol biosynthesis are designated according to WikiPathways (https://www.wikipathways.org/pathways/WP4718.html), and their order corresponds to the Bloch pathway. SREBF2, sterol regulatory element-binding transcription factor 2; INSIG, insulin induced gene; SCAP, SREBP cleavage-activating protein; SREBP, sterol regulatory element-binding protein; HMGCR, 3-hydroxy-3-methylglutaryl-coenzyme A reductase; MVK, mevalonate kinase; FDPS, farnesyl diphosphate synthase; FDFT1, farnesyl-diphosphate farnesyltransferase 1; SS, squalene synthase; SQLE, squalene epoxidase; CYP51A1, cytochrome P450 family 51 subfamily A member 1; TM7SF2, transmembrane 7 superfamily member 2; HSD17B7, hydroxysteroid 17-beta dehydrogenase 7; EBP, EBP cholestenol delta-isomerase; DHCR7, 7-dehydrocholesterol reductase; DHCR24, 24-dehydrocholesterol reductase; ACAT2, acetyl-CoA acetyltransferase 2; HMGCS, 3-hydroxy-3-methylglutaryl-CoA synthase 1; PMVK, phosphomevalonate kinase; MVD, mevalonate diphosphate decarboxylase; IDI, isopentenyl-diphosphate delta isomerase; GGPS1, geranylgeranyl diphosphate synthase 1; LSS, lanosterol synthase; LBR, lamin B receptor; MSMO1, methylsterol monooxygenase 1; NSDHL, NAD(P) dependent 3-beta-hydroxysteroid dehydrogenase NSDHL; SC5D, sterol-C5-desaturase.

2. Materials and Methods

2.1 Collecting Data on SREBP-Binding Sites

Data on SREBP-binding sites were extracted manually from research articles found in TRRD (https://wwwmgs.bionet.nsc.ru/mgs/gnw/trrd/), TRRUST (https://www.grnpedia.org/trrust/), and PubMed (https://pubmed.ncbi.nlm.nih.gov/) [39,40]. The articles available at PubMed in August 2025 were reviewed. The query terms that were used are listed in **Supplementary Table 1**. Among the articles found in this way, we selected manually those that contained the following data on SREBP binding sites: (1) nucleotide sequences, positions relative to transcription start site (TSS); (2) descriptions of experiments confirming specific binding of sites to SREBP using electrophoretic mobility shift assay and experiments confirming functional activity of sites based on the transient transfection technique. We took into consideration SREBP-binding sites in the regulatory regions of the genes from the core part of the gene network we had built [16]. DNA sequences with a length of 30 nucleotides were extracted from the research articles, including the SREBP-binding motifs (10 nucleotides) and its flanking regions. The genomic positions of these

30-nucleotide sequences were identified using the Blast-Like Alignment Tool (BLAT, https://genome.ucsc.edu/cgi-bin/hgBlat) provided by the UCSC Genome Browser (https://genome.ucsc.edu). BLAT allows rapid search for sequences of 95% and greater similarity and of 25 bases or longer in a single target genome or in all available genomes. BLAT also determines positions of the query sequence in the target genome [41]. For every single query sequence (30-nucleotide DNA sequence including a SREBP site) the following query parameters were used: "Genome: human (hg38)", "Assembly: Dec. 2013 (GRCh38/hg38)", "Query type: BLAT's guess", "Sort output: Query,score", "Output type: Hyperlink", "All Results (no minimum matches)".

2.2 Finding Genetic Variants Affecting SREBP-Binding Sites

The set of genetic variants colocalized with SREBP-binding sites was obtained from dbSNP155 using the UCSC Table Browser data retrieval tool [42]. The following query parameters were used: "Assembly:Dec. 2013 (GRCh38/hg38)", "Group: Variation", "Track: dbSNP155", "Table: All dbSNP (155) (dbSnp155)", "Define region of interest: Define regions". Region definitions (genomic positions of 22 SREBP-binding sites) were pasted in a three-field bed file format (example line: "chr12 109573589 109573618"). Then the following options were selected: "Output format: All fields from selected table", "Output field separator: tsv (tab-separated)", "File type returned: Plain text", "Get output". We extracted deletions whose positions overlapped with SREBP sites and SNVs colocalized with these sites or their three 5′- or 3′-flanking nucleotides.

2.3 Ranking Genetic Variants Based on Data on the Evolutionary Conservation of Nucleotides

The evolutionary conservation of the nucleotides affected by genetic variants was quantified using the PhyloP score [43,44]. The PhyloP score values for each nucleotide were obtained using the UCSC Table Browser data retrieval tool [42]. PhyloP score evaluates signatures of selection at particular nucleotides based on multiple alignments of homologous regions of the genomes of 100 vertebrate species. PhyloP first estimates the mean number of substitutions in each genome to evaluate the neutral evolution rate. Then the probability of observed substitutions is analyzed on the assumption of neutral evolution. Sites predicted to be conserved are assigned positive scores, whereas sites predicted to be fast evolving are assigned negative scores. The absolute values of the PhyloP score represent the -log (*p*-value) under the null hypothesis of neutral evolution [43,45].

2.4 Quantitation of SREBP-Binding Site Similarity to SREBP-Binding Motif Models

To predict the binding preference of each SREBP site from our collection and to determine the potential impact of genetic variants on the site functional activity we used MoloTool (https://molotool.autosome.org/). This program [46] performs sequence scanning to identify transcription factor binding motif occurrences in a given set of DNA sequences and quantifies the similarity of DNA sequences (*p*-value) to the corresponding site models (matrices) from the HOCOMOCO database (https://hocomoco14.autosome.org/) [24]. The scanning was performed at *p*-value cutoff = 5×10^{-2} using the SRBP1.H14CORE.0.P.B and SRBP2.H14CORE.1.P.B models. According to the HOCOMOCO database documentation, the motif models SRBP1.H14CORE.0.P.B and SRBP2.H14CORE.1.P.B were derived from ChIP-Seq data. In this way, *p*-value1 and *p*-value2 were obtained, characterizing the similarity of the DNA sequences to the SRBP1.H14CORE.0.P.B and SRBP2.H14CORE.1.P.B matrices, respectively.

The sites were designated as SREBP1- or SREBP2-preferring if the *p*-value1 / *p*-value2 ratio was less than 0.5 or greater than 2, respectively. If the *p*-value1 / *p*-value2 ratio was between 0.5 and 2, the site was classified as a binding site for both factors.

To determine the potential impact of a genetic variant on the site functional activity for both SREBP motif models, the *p*-value (majorAllele) and *p*-value (minorAllele) were determined. The genetic variant was considered damaging if the *p*-value (majorAllele) / *p*-value (minorAllele) ratio was less than 1.

3. Results

3.1 The Set of SREBP-Binding Sites Identified Experimentally

Analyzing the scientific publications manually, we collected data on 22 SREBP-binding sites, which were identified experimentally (Table 1). These sites were found within the regulatory regions of 12 genes, 11 of which encode enzymes involved in cholesterol biosynthesis and one gene (*SREBF2*) encodes a transcription factor.

According to evaluations of similarity with SREBP site models from the HOCOMOCO database (see Subsection 2.4), more than half of the sites (13 of 22, or 59%) were found to have higher similarity to the SREBP2-binding motif model (**Supplementary Table 2**) and were classified as SREBP2-preferring sites (Table 1, rightmost column). This observation is consistent with the frequently stated thesis that SREBP2 is the main regulator of cholesterol biosynthesis [26,47]. Five of 22 sites were classified as SREBP1-preferring sites, and four sites were classified as sites that bind both factors equally.

3.2 Genetic Variants Colocalized With SREBP-Binding Sites

Comparing the genomic coordinates of 22 SREBP-binding sites (Table 1) with positions of genetic variants presented in the dbSNP155 database, we identified variants colocalized with the core or flanking DNA sequences of

Table 1. SREBP-binding sites in the regulatory regions of genes from the gene network controlling cholesterol biosynthesis extracted from scientific publications.

DNA sequences of the SREBP-binding site and nearest flanking regions *	Gene **	Molecular function of the encoded protein	Position of the SREBP site relative to the transcription start site of the gene ***	Reference (PMID)****	Genomic positions of the 30-nucleotide sequence in GRCh38/hg38 *****	Predicted binding preference of the site ******
tggccgagccGTGGTGAGAGatggtgcggt	*HMGCR*	Enzyme, 3-hydroxy-3-methyl-glutaryl-coenzyme A reductase (converts 3-hydroxy-3-methyl-glutaryl-CoA to mevalonate)	–180/–171	7592765	chr5:75,337,024–75,337,053	SREBP2-preferring
gtggtgagagATGGTGCGGTgcctgttctt	*HMGCR*		–170/–161	7592765	chr5:75,337,034–75,337,063	SREBP2-preferring
ttcgcagttctCACCCCAGggcgggacact	*MVK*	Enzyme, mevalonate kinase (converts mevalonic acid to 5-phosphomevalonate)	–288/–281	9464248	chr12:109,573,589–109,573,618	SREBP2-preferring
gtgcgccatgCTCACACGACggccagcagc	*FDPS*	Enzyme, farnesyl pyrophosphate synthase (converts isopentenyl diphosphate + dimethylallyl diphosphate to geranyl diphosphate)	–234/–225	20450493	chr1:155,308,715–155,308,744	SREBP1-preferring
ctagagtgttATCACGCCAGtctccttccg	*FDFT1/SS*	Enzyme, squalene synthase (converts farnesyl diphosphate to squalene)	–189/–180	7665618	chr8:11,802,538–11,802,567	SREBP2-preferring
cccgtcagtcCACCCCACgaggccgcag	*FDFT1/SS*		–108/–101	7665618	chr8:11,802,619–11,802,646	SREBP2-preferring
cggggtcttcCTAGTGTGAGcggccctggc	*FDFT1/SS*		–145/–136	11008488	chr8:11,802,580–11,802,609	Both factors
atagcgccatCTCGGCCTACcgcgcggtgc	*SQLE*	Enzyme, squalene monooxygenase (converts squalene to 2,3-epoxysqualene)	–185/–176	28342963	chr8:124,998,342–124,998,371	SREBP2-preferring
tcggcctaccGCGCGGTGCGcgccctcatt	*SQLE*		–174/–165	28342963	chr8:124,998,353–124,998,382	Both factors
ccaacccgcaCCACGCAACatcgccctag	*SQLE*		–261/–253	12083769	chr8:124,998,257–124,998,285	SREBP2-preferring
gtaggccgagATCACCTCAGgcgctcgcgg	*CYP51A1*	Enzyme, lanosterol 14-alpha demethylase (converts lanosterol to 14-demethyllanosterol)	–232/–223	10551787	chr7:92,134,811–92,134,840	SREBP1-preferring
gccgtgtcccGGCACCCCGCacccgggcac	*CYP51A1*		–135/–126	10551787	chr7:92,134,714–92,134,743	Both factors
gaggcgcccgCGTCGGGTCCGcccccgagcg	*TM7SF2*	Enzyme, delta(14)-sterol reductase (converts 4,4-dimethylcholesta-8,14,24-trienol to 14-demethyllanosterol)	–241/–231	16784888	chr11:65,111,764–65,111,794	SREBP1-preferring

Table 1. Continued.

DNA sequences of the SREBP-binding site and nearest flanking regions *	Gene **	Molecular function of the encoded protein	Position of the SREBP site relative to the transcription start site of the gene ***	Reference (PMID)****	Genomic positions of the 30-nucleotide sequence in GRCh38/hg38 *****	Predicted binding preference of the site ******
ctcggcctatcTCACCCGAcgccgtctccg	*HSD17B7*	Enzyme, 17-beta-hydroxysteroid dehydrogenase 7 (converts 14-demethyllanosterol to zymosterol)	–197/–190	15862973	chr1:162,790,593–162,790,622	SREBP1-preferring
gtcccagtctCCACCCCAAacctaactca	*EBP*	Enzyme, 3-beta-hydroxyster-oidDelta(8),Delta(7)-isomerase (converts zymosterol to 5alpha-cholesta-7,24-dien-3beta-ol)	–165/–157	12855700	chrX:48,521,655–48,521,683	SREBP2-preferring
caggtgaaaaTGGCGCCACGcgcgggggcg	*EBP*		–138/–129	12855700	chrX:48,521,682–48,521,711	SREBP2-preferring
gttgggagcgGCACGGGCCgccttcttcg	*EBP*		–78/–70	12855700	chrX:48,521,742–48,521,770	Both factors
tgggcccggaGTCAGCTCAGgctcctcgac	*DHCR7*	Enzyme, 7-dehydrocholesterol reductase (converts 7-dehydro-cholesterol to cholesterol)	–155/–146	25048193	chr11:71,448,487–71,448,516	SREBP1-preferring
tgacatcatcCGGGGGCGGGcgcccctgcc	*DHCR7*		–55/–46	25048193	chr11:71,448,396–71,448,425	SREBP2-preferring
ctgcgggttcCTGGTCCGATccccggcgcg	*DHCR24*	Enzyme, delta(24)-sterol reductase (converts desmosterol to cholesterol)	–205/–196	22809995	chr1:54,887,305–54,887,334	SREBP2-preferring
ccgcccgggtCTCGGCCCCACcgaacctcgg	*DHCR24*		–169/–160	22809995	chr1:54,887,260–54,887,289	SREBP2-preferring
atgacgcgccATCACCCCACgcaccgcttc	*SREBF2*	Transcription factor, sterol regulatory element-binding protein 2	–128/–119	8900111	chr22:41,832,961–41,832,990	SREBP2-preferring

* SREBP-binding motif (core DNA sequence) is shown in capital letters.

** The order of genes corresponds to the order of enzymes in the cholesterol biosynthesis pathway that is given in WikiPathways diagram (https://www.wikipathways.org/pathways/WP4718.html). Starting with CYP51A1, the order in which the genes are listed in the table corresponds to the Bloch pathway.

*** As given in the research article dedicated to the experimental identification of this site.

**** According to these publications, SREBP sites were studied in an electrophoretic mobility shift assay and in a transient transfection assay.

***** The identification of genomic positions of the SREBP sites is described in Subsection 2.1.

****** The predicted binding preference for a given site is indicated based on *p*-values calculated by MoloTool (https://molotool.autosome.org/), see details in Subsection 2.4. and Subsection 3.1.

Table 2. Deletions colocalized with SREBP-binding sites and 20 SNVs colocalized with SREBP-binding sites that have the highest ranks according to PhyloP scores of affected nucleotides*.

Gene	Number of SREBP -binding sites **	Reference (PMID)	Total number of SNVs ***	Del ****	20 SNVs with the highest ranks according to PhyloP scores *****
HMGCR	2	7592765, 28342963	15		rs1347256714, rs1759792342, rs1759792538, rs1759793207
MVK	1	9464248	7	rs1264806290	no
FDPS	1	20450493	1		no
FDFT1/SS	3	7665618, 11008488	35	rs1563290033, rs1585844469, rs1029411034	**rs1680036166**, **rs1357951184**, rs1308720772, rs1806253588, rs1806253730, rs750930727
SQLE	3	28342963, 12083769	24		**rs1461664179**, **rs1814774941**, ***rs1167673458***, *rs1351295252*, *rs899347017*, rs954597419
CYP51A1	2	10551787, 12145339	16		no
TM7SF2	1	16784888	8		no
HSD17B7	1	15862973	8		no
EBP	3	12855700	16	rs1207705310	no
DHCR7	2	25048193	18		no
DHCR24	2	22809995	15		*rs1647107097*
SREBF2	1	8900111	8		*rs1444526074*, *rs1293790327*, rs1602252939
Total number	22		171	5	20

* Genes are listed in the order corresponding to the order of encoded enzymes operating in the cholesterol biosynthesis pathway.

** The functional activity of these sites was confirmed by transient transfection assay and electrophoretic mobility shift assay.

*** The total number of SNVs colocalized with SREBP-binding sites or their nearest three left and right flanking nucleotides (The full list of these SNVs is presented in **Supplementary Table 4**).

**** The 16-nucleotide-long deletion noted by UCSC Table Browser data retrieval tool as ClinvarPathogenic is underlined. All other deletions result in the loss of one nucleotide.

***** 20 SNVs affecting nucleotides with the highest PhyloP scores (between 4.29838 and 6.11165). Five SNVs with PhyloP score >5 are typed in bold; six SNVs colocalized with flanking nucleotides are typed in italics.

SNVs, single-nucleotide variants; PhyloP, phylogenetic *p*-values; UCSC, University of California, Santa Cruz.

SREBP-binding sites. Among them, there were the following two groups of variants that will be considered in this paper: (1) five deletions affecting the central SREBP-binding motif (Table 2, the fifth column; **Supplementary Table 3**); (2) 171 SNVs, colocalized with the central 10-nucleotide SREBP-binding motif or the nearest three left or three right flanking nucleotides (**Supplementary Table 4**).

3.3 Deletions Affecting SREBP-Binding Sites

As mentioned above, five deletions whose positions overlap with the positions of the analyzed SREBP-binding sites were found among the genetic variants (Table 2 and **Supplementary Table 3**). These deletions were identified within regulatory regions of three genes: mevalonate kinase (*MVK*) farnesyl-diphosphate farnesyltransferase 1 (*FDFT1*), and *EBP cholestenol delta-isomerase* (*EBP*). All variants of this group are rare: their minor allele frequencies in the studied datasets ranged from 3.78×10^{-6} to 0.00274725 (**Supplementary Table 3**, fifth column).

It is noteworthy that rs1563290033, which is a 16-nucleotide deletion, is characterized in the ClinVar database as Pathogenic. As shown in Fig. 2, the DNA sequence corresponding to variant rs1563290033 includes the SREBP-binding site that has been identified experimentally in the regulatory region of the *FDFT1* gene [48]. The genomic positions of this 10-nucleotide site identified by Tansey and Shechter [48] are chr8:11,802,590–11,802,599.

The other four deletions (rs1585844469, rs1029411034, rs1207705310, and rs1264806290) resulted in the loss of one nucleotide. None of them was found in the genome-wide association study (GWAS) catalog (https://www.ebi.ac.uk/gwas/) as associated with any cholesterol-related trait. Two 1-bp deletions (rs1585844469 and rs1029411034) affected the same

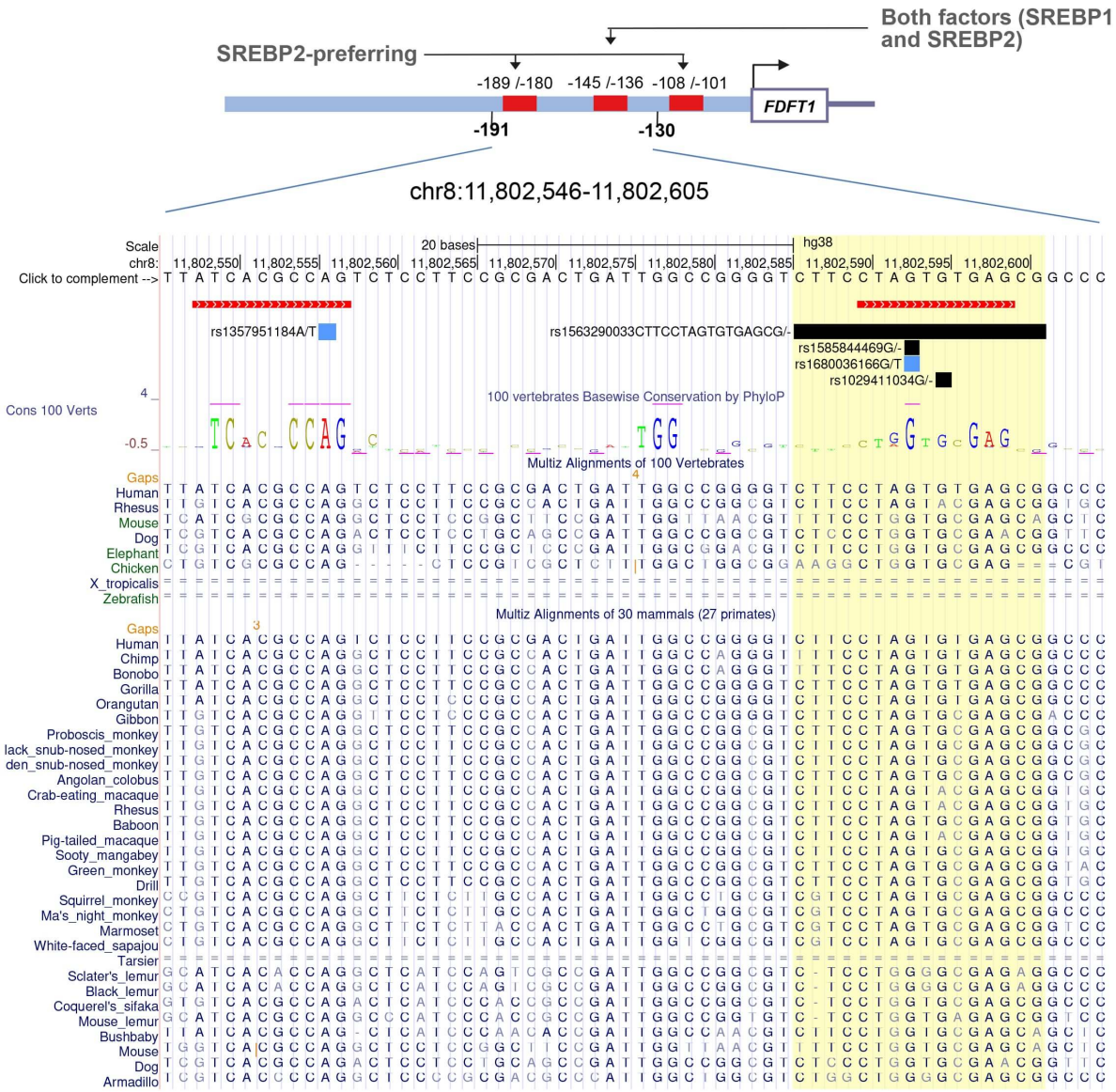


Fig. 2. Fragment of the human *FDFT1* regulatory region harboring functionally active SREBP-binding sites (shown in red bars). For a detailed description of these SREBP-binding sites, see Table 1. SREBP-binding sites were predicted to be SREBP2-preferring or interacting with both factors (SREBP1 and SREBP2) using MoloTool (see Subsection 2.4 and **Supplementary Table 2**). The graphical representation provided by the UCSC Genome Browser includes the following tracks (1) the nucleotide sequence (chr8:11,802,546–11,802,605) of the *FDFT1* regulatory region; (2) SREBP-binding sites (custom track); (3) genetic variants affecting SREBP-binding sites: (i) deletions (indicated with black bars) rs1563290033 (loss of 16 nucleotides), rs1585844469 (loss of one nucleotide), and rs1029411034 (loss of one nucleotide), as well as (ii) two SNVs (shown in blue bars): rs1357951184 and rs1680036166; (4) basewise conservation in 100 vertebrates according to PhyloP; (5) Multiz alignments of 100 vertebrates; (6) Multiz alignments of 30 mammals.

SREBP-binding site in the *FDFT1* promoter region as the 16-nucleotide deletion rs1563290033.

We quantified the impact of the four above-mentioned 1-bp deletions on the similarity to site models. Previously, sites were classified according to their preference for the factor (Table 1, **Supplementary Table 2**). The SREBP site at –288/–281 in the *MVK* promoter was classified as SREBP2-preferring and the other two SREBP sites (–145/–136 in *FDFT1* promoter and –78/–70 in *EBP* promoter) were classified as sites equally similar to the models of both factors. The analysis performed using MoloTool and SREBP1 and SREBP2 motif models (matrices) from the HOCOMOCO database showed that all four 1-bp deletions decreased the similarity of the SREBP sites to both motif models (**Supplementary Table 5**). Thus, it can be expected that all four deletions reduce the functional activity of the SREBP- binding sites.

3.4 Ranking of 171 SNVs Located Within SREBP-Binding Sites or Colocalized With the Nearest Three Left or Right Flanking Nucleotides Based on the Evolutionary Conservation of Affected Nucleotides

These 171 *SNVs* were found within the regulatory regions of all 12 genes for which we collected information (Table 1, **Supplementary Table 4**). ClinVar contains no information on any of these variants. None of the single genetic variants that we revealed was annotated in the GWAS catalog (https://www.ebi.ac.uk/gwas/) as associated with any cholesterol-related trait. Using the PhyloP score metrics (see Subsection 2.3), we quantified the evolutionary conservation of the nucleotides affected by the SNVs identified in the study (**Supplementary Table 4**, fourth column) and ranked SNVs based on these metrics. Positive PhyloP scores (starting from 0.0229606 and ending with 6.11165) were found for 126 out of 171 SNVs, PhyloP scores more than 2 were found for 50 SNVs. It was found that 20 SNVs affect rather conservative nucleotides: PhyloP score ranging from 4.2984 to 6.1117 (Table 2, **Supplementary Table 4**). These 20 SNVs (**Supplementary Table 6**) were found in the regulatory regions of four enzyme-encoding genes (*HMGCR*, *FDFT1*, *SQLE*, and 24-dehydrocholesterol reductase (*DHCR24)*) as well as in the regulatory region of *SREBF2*, encoding a transcription factor. Fourteen of these 20 SNVs affect nucleotides within the SREBP-binding motif, and six (italicized in Table 2) affect flanking nucleotides. According to UCSC table browser data, all 20 SNVs are classified as rare: their minor allele frequencies in the studied population samples range from 3.78×10^{-6} to 0.0021575 (**Supplementary Table 6**, eighth column).

Five of these 20 SNVs replaced the most conserved nucleotides, whose PhyloP scores exceeded 5. Two variants (rs1461664179 and rs1814774941) were found within the SREBP2-preferring site located in the *SQLE* promoter. The other two variants (rs1680036166 and rs1357951184) were found in the *FDFT1* promoter within the SREBP site that was predicted to bind both factors equally. Finally, rs1167673458 affected a nucleotide flanking the SREBP2-preferring site in the *SQLE* promoter.

It is worth noting that the other three of 20 SNVs from **Supplementary Table 6** were located within or near the SREBP site in the promoter of the *SREBF2* gene. One of these three SNVs (rs1602252939) affected a nucleotide with PhyloP score = 4.65191 within the SREBP site and two SNVs (rs1444526074, rs1293790327) affected flanking nucleotides.

Using MoloTool we quantified the impact of the five above-mentioned variants localized in SREBP sites (rs1461664179, rs1680036166, rs1814774941, rs1357951184, and rs1602252939) on the similarity to SREBP-binding motif models. The initial analysis performed for the reference alleles showed that three of the four sites where the variants in question were found are SREBP2-preferring, and only the –145/–136 site in the *FDFT1* promoter was classified as a site equally similar to the models of both factors (**Supplementary Table 2** and **Supplementary Table 5**). The minimum *p*-value reflecting the similarity of the site to the SREBP2-binding motif model was 8.17×10^{-8} (SREBP-binding site at –128/–119 in the *SREBF2* promoter), and the maximum *p*-value was 7.52×10^{-3} (SREBP-binding site at –145/–136 in the *FDFT1* promoter) (**Supplementary Table 2**). It was found (**Supplementary Table 5**) that three SNVs (rs1680036166 and rs1357951184 (in *FDFT1* promoter), and rs1602252939 (in *SREBF2* promoter)) decreased similarity to both motif models (SREBP1 and SREBP2). Based on this result, we hypothesize that these three SNVs disrupt the functional activity of SREBP sites with a high degree of probability. In addition, SNV rs1814774941 (–185/–176 site in the *SQLE* promoter) decreased similarity to the SREBP2-binding motif model but did not affect the similarity of this SREBP2-preferring site to the SREBP1-binding motif model. Therefore, we assume that rs1814774941 may also be damaging.

4. Discussion

4.1 Genetic Variants Altering Transcription Factor Binding Sites

Transcription factors from the SREBP subfamily provide transcriptional control of target genes (including genes encoding enzymes of the cholesterol biosynthesis pathway) according to the intracellular cholesterol level [17,49]. This mechanism is effected by specific DNA-protein interactions between SREBPs and their binding sites on DNA [50,51]. The specific binding of all transcription factors to DNA is driven by a certain linear sequence of nucleotides within binding sites. The nucleotide sequence defines conformational and physicochemical properties of DNA, thus determining the three-dimensional DNA structure optimal for interaction with a particular protein [52]. Within each binding site, there are nucleotides that interact with the transcription factor most tightly [50]. These nucleotides tend to be quite conserved both in evolutionary terms and in the alignments of binding sites for an individual transcription factor within a single genome [53,54,55].

The nucleotide sequence of a binding site may be altered by mutations [56,57,58]. An increasing number of studies are dedicated to the link between genetic variants in the regulatory regions of genes affecting transcription factor binding sites and the expression levels of these genes [59,60,61,62]. Many of these variants are pathogenic that is, they are associated with predisposition to diseases [37,38,63].

Since it is known that decreased activity of cholesterol biosynthesis enzymes due to mutations in the coding regions of the corresponding genes is accompanied by morbid conditions [34,35,36,64], it is reasonable to expect that decreased expression of these genes can also cause such states. Therefore, the aim of our study was to identify ge-

netic variants that could damage SREBP-binding sites and lead to decreased expression of the corresponding gene at low intracellular cholesterol levels.

In our work, we sought and analyzed genetic variants within stretches of 16 nucleotides, covering SREBP sites discovered experimentally. The choice of the 16-nucleotide DNA fragment was motivated by the following considerations: (1) the 2.3 Å resolution of the cocrystal structure of the SREBP-1a homodimer bound to the DNA fragment derived from the human LDL receptor gene promoter had demonstrated that SREBP-1a comes into contact with its recognition site on DNA, which is 10 nucleotides long [50]; (2) HOCOMOCO database [24] presents SREBP1- and SREBP2-binding motif models derived from the ChIP-Seq data that are 12 nucleotides long. Thus, the 16 nucleotides considered in this study include both nucleotides involved in DNA-protein contacts and all adjacent nucleotides that match SREBP1- or SREBP2-binding motif patterns derived from ChIP-Seq and therefore may be important for binding.

4.2 Deletions Colocalized With SREBP-Binding Sites

We identified five deletions that are colocalized with SREBP-binding sites in the regulatory regions of the genes encoding enzymes participating in cholesterol biosynthesis (Table 2, and **Supplementary Table 3**). It was found that rs1563290033 is characterized in ClinVar as Pathogenic, and it leads to the loss of the SREBP site from the *FDFT1* regulatory region (Fig. 2). The rs1563290033 variant is a deletion of 16 nucleotides, 10 of which correspond to the proximal SREBP-binding site near *FDFT1*, as well as 4 and 2 nucleotides flanking this site.

The clinical significance of rs1563290033 was established by Coman D. and coauthors, who studied the genetic basis of squalene synthase (SS) deficiency [64]. SS is an enzyme of the cholesterol biosynthesis pathway encoded by the *FDFT1* gene. It was found that one individual (ethnically European) with squalene synthase deficiency had a 16-nucleotide deletion (Chr8(GRCh37):g.11660095_11660110del) located in an *FDFT1* intronic region, and this deletion matched the positions of rs1563290033. According to [64], the clinical manifestations of the deletion include seizures, irritability, dysmorphism, fixed flexion deformities in elbows, hypospadias, poor sleep initiation and maintenance, decreased plasma cholesterol.

According to the Ensembl genome database, *FDFT1* has several transcripts. For some of them, including ENST00000220584.9, classified as *MANE Select*, the rs1563290033 deletion disrupts the nucleotide sequence of the promoter region. To test the hypothesis that the deleted 16-bp DNA fragment has a regulatory activity, Coman D. and coauthors performed a luciferase assay using a reporter construct that included the promoter with the 16-bp deletion. It turned out that this deletion impaired the activity of the promoter and lowered the *FDFT1* transcription level compared to the control. However, the ability of the 16-nucleotide DNA sequence to interact with SREBPs was not tested in that study. We found that rs1563290033 colocalizes with the SREBP-binding site, which hints at a mechanism that reduces the *FDFT1* promoter activity.

We also identified four single-nucleotide deletions (rs1585844469, rs1029411034, rs1207705310, and rs1264806290) that affected SREBP-binding sites upstream of three genes: *FDFT1*, *EBP*, *MVK* (Table 2 and **Supplementary Table 3**). According to the analysis performed using MoloTool, all four single-nucleotide deletions reduce the similarity to the models of both binding sites: for SREBP1 and for SREBP2 (**Supplementary Table 5**). Thus, we conjecture that these four single-nucleotide deletions disrupt the functional activity of SREBP sites by causing a frameshift in the binding site motif sequence. This shift could alter helical phasing and recognition of the entire adjacent DNA element, affecting the regulation of gene expression.

4.3 Top-Ranked SNVs Colocalized With SREBP-Binding Sites and Notable Genes

To identify SNVs that may influence the expression levels of the genes in question, we ranked 171 SNVs (**Supplementary Table 4**) located within SREBP sites or in the closest vicinity (3 bp to the left or to the right) according to PhyloP scores of affected nucleotides. Four SNVs (in bold in Table 2), affecting nucleotides within SREBP sites had the highest PhyloP scores (more than 5.0), which means that these four SNVs affected nucleotides characterized by the highest evolutionary conservation. These four SNVs are: (1) rs1814774941 and rs1461664179 (upstream of *SQLE*); (2) rs1680036166 and rs1357951184 (upstream of *FDFT1*). According to the analysis performed using MoloTool, all of these four SNVs reduce the similarity to the SREBP2 site model, and two of the four (rs1680036166 and rs1357951184) reduce the similarity to the SREBP1 site model.

Comparing the results based on ranking by PhyloP score with the data on deletions, we found that the most interesting genetic variants are concentrated in SREBP-binding sites located within promoter regions of two genes: *FDFT1* and *SQLE*.

SREBP-binding sites from the *FDFT1* promoter region [48,65] may be altered by two SNVs (rs1680036166 and rs1357951184) that had extremely high PhyloP scores (more than 5.0) as well as by three deletions: (1) rs1563290033, a deletion of 16 nucleotides discussed in detail above; and (2) two single-nucleotide deletions, rs1585844469 and rs1029411034 (Fig. 2). According to MoloTool, both SNVs and both 1-bp deletions decrease the similarity to SREBP-binding motif models (**Supplementary Table 5**). The genetic variants named here are rare. The rs1680036166 SNV, affecting a nucleotide with PhyloP score = 5.87398, was found in

the allele frequency panel of 3552 Japanese individuals 3.5KJPNv2 [66] with a frequency of 0.00004. The rs1357951184 SNV, affecting nucleotide with PhyloP score = 5.41036 was found in the TOPMED dataset, including about 158,000 individuals of different ethnicities [https://topmed.nhlbi.nih.gov], with a frequency of 7.55601×10^{-6}, and in the gnomAD genome dataset, including samples from 41,462 Africans [https://gnomad.broadinstitute.org/], with a frequency of 0.00002. The rs1585844469 1-bp deletion was found in the allele frequency panel of 3552 Japanese individuals 3.5KJPNv2 [66] with a frequency of 0.00178998 and in the Korea1K dataset, containing 1094 Korean personal genomes [67], with a frequency of 0.00274725. The other 1-bp deletion (rs1029411034) was found in TOPMED with a frequency of 3.778×10^{-6}.

One of the two SREBP sites within the *SQLE* promoter region [68] is also enriched in top-ranked SNVs (Fig. 3). There are two SNVs (rs1814774941 and rs1461664179) colocalized with this site and SNV rs1167673458, altering the second nucleotide on the left flank. All three SNVs affect highly evolutionarily conserved nucleotides (their PhyloP scores are 5.76627, 6.11165, and 5.67992, respectively). In addition, rs1814774941 was predicted to decrease the similarity of the SREBP2-preferring site to the SREBP2-binding motif model without changing the similarity to the SREBP1 binding site model (**Supplementary Table 5**). The substitution rs1814774941 is rare. It was found in TOPMED with a frequency of 3.778×10^{-6}.

It is remarkable that the *SQLE* and *FDFT1* genes, whose promoter regions contain genetic variants identified among the five SNVs with the highest ranks, encode enzymes (squalene epoxidase and squalene synthase, respectively) that carry out two consecutive reactions in the cholesterol biosynthesis pathway (for details, see Fig. 1).

It is noteworthy that among the 20 SNVs affecting highly evolutionarily conserved nucleotides (the sixth column in Table 2 and **Supplementary Table 6**), there is rs1602252939 (PhyloP score = 4.65191) located in the SREBP site upstream of *SREBF2*. *SREBF2* encodes a transcription factor that is a key regulator of cholesterol biosynthesis [25]. The rs1602252939 SNV was found in the Korea1K dataset [67] with a frequency of 0.00163755.

Thus, by ranking SNVs by PhyloP score which characterizes the evolutionary conservation of nucleotides in the human genome, and quantifying the similarity of the corresponding SREBP sites to the two SREBP-binding motif models, we identified variants that may affect the activity of SREBP-binding sites with a high probability. Three of the variants considered above were found in datasets containing Korean or Japanese personal genomes. These variants are (1) SNV (rs1680036166) in the *FDFT1* promoter, identified among the five SNVs with the highest ranks (in bold in the sixth column of Table 2, and **Supplementary Table 6**); (2) SNV (rs1602252939) in the SREBP-binding site located in the *SREBF2* promoter (identified among the 20 top-ranked SNVs (the sixth column in Table 2 and **Supplementary Table 6**)); (3) the deletion of one nucleotide (rs1585844469) in the *FDFT1* promoter (**Supplementary Table 3**). Perhaps such co-occurrence is not mere coincidence; it may reflect a common origin of these ethnic groups. As noted earlier, squalene synthase deficiency is a rare disorder, so far identified only in Europeans [64]. On the other hand, the variant rs1585844469 has an allele frequency of ~0.0018–0.0027 in the Japanese and Korean populations. Therefore, it is reasonable to assume that the manifestation of this disorder in individuals of Korean and Japanese ethnicities who are homozygous for rs1585844469 may be attenuated by a combination of ethnic-specific genetic and environmental factors.

4.4 Study Limitations

This study has several limitations. (1) We focused here only on SNVs and short deletions and did not analyze the effects of insertions. This is because: (i) PhyloP score metrics are not applicable to nucleotides within insertions; (ii) *in silico* quantitation of the impact of insertions on site functionality is highly problematic. The DNA sequence corresponding to the alternative allele may have greater similarity to the site model; however, the insertion may change the position of the site relative to other regulatory elements within the promoter region, which may oppose the positive effect of such insertion on the contextual features of the site. (2) In the case of a deletion, reliable prediction of the impairment of the transcription factor binding site *in silico* is only possible if the genetic variant results in the complete deletion of the site, as with the rs1563290033 16-base pair deletion. Therefore, when considering all other genetic variants, we keep in mind a potential damaging effect. (3) The promoter region of a gene may contain multiple sites, so that the disruption of one site may only have a minor effect on expression levels. (4) Finally, the phenotypic manifestations of diminished SREBP site activity may be attenuated by a combination of ethnic-specific genetic and environmental factors.

5. Conclusions

In this study, we combined data on experimentally confirmed SREBP-binding sites, which we extracted manually from publications; data on genetic variants from dbSNP; genome-wide data on the evolutionary conservation of each nucleotide in the human genome; and *in silico* estimations of the effects of genetic variants on the similarity of SREBP sites to SREBP-binding motif models.

We considered SREBP-binding sites in the promoter regions of the genes involved in the gene network regulating cholesterol biosynthesis. Disturbances in cholesterol biosynthesis lead to health problems: significant birth malformations in early ontogenesis and putative increased risk of metabolic, liver, and cardiovascular diseases in adults. We found genetic variants colocalized with SREBP-

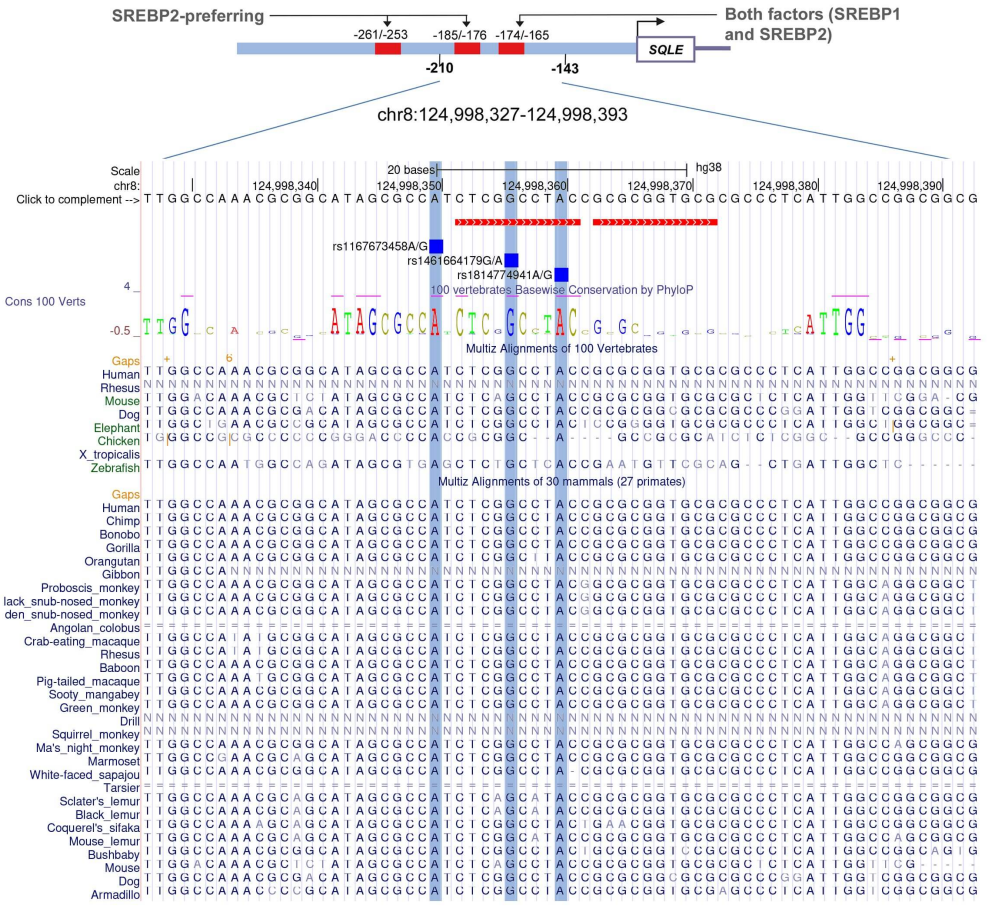


Fig. 3. Fragment of the human *SQLE* regulatory region harboring functionally active SREBP-binding sites (indicated with red bars). For a detailed description of these SREBP-binding sites, see Table 1. The SREBP-binding sites were predicted to be SREBP2-preferring or interacting with both factors (SREBP1 and SREBP2) using MoloTool (see Subsection 2.4 and **Supplementary Table 2**). The graphical representation provided by the UCSC Genome Browser includes the following tracks (1) the nucleotide sequence (chr8:124,998,327–124,998,393) of the *SQLE* regulatory region; (2) SREBP-binding sites (custom track); (3) SNVs affecting SREBP site or its flanking nucleotide (shown with blue bars): rs1461664179, rs1814774941, rs1167673458; (4) 100 vertebrates basewise conservation by PhyloP; (5) Multiz alignments of 100 vertebrates; (6) Multiz alignments of 30 mammals.

binding sites: five deletions (Table 2 and **Supplementary Table 3**) that disrupt SREBP sites, as well as 171 SNVs (**Supplementary Table 4**).

When ranking SNVs according to the evolutionary conservation of the nucleotides they affect, we found that the variants affecting the most conserved nucleotides are located in the *FDFT1*, *SQLE* and *SREBF2* promoters. We also detected four single-nucleotide deletions altering SREBP-binding sites in the *FDFT1*, *EBP*, and *MVK* promoter regions.

Subsequent computer analysis aimed at quantifying the effect of variants on the similarity of the sites in question to SREBP1 and SREBP2 binding motif models showed that three SNVs (rs1680036166, rs1357951184, and rs1602252939) and four 1-bp deletions (rs1585844469, rs1029411034, rs1207705310, and rs1264806290) are highly likely to affect the binding capacity and functional activity of the SREBP sites. Further experimental studies using the electrophoretic mobility shift assay and the reporter assay are needed to confirm this idea and to classify these variants as clinically significant. In any case, such probability should be taken into account when assessing the risk of pathologies in individuals carrying alternative alleles of these genetic variants.

Among the five deletions we identified as colocalized with SREBP sites, there was rs1563290033 in the *FDFT1* promoter. It is known as pathogenic (clinically significant), associated with squalene synthase deficiency. The comparison of genomic positions of the rs1563290033 16-bp deletion with the positions of the SREBP-binding site allowed us to propose a mechanism for its pathological effect: the decrease in *FDFT1* expression is caused by the loss of the SREBP-binding site.

The approach we used to seek potentially clinically significant variants in the regulatory regions of genes can

be applied in future to the search for such variants upstream of genes from other gene networks.

Abbreviations

FDFT1, farnesyl-diphosphate farnesyltransferase 1; SNV, single-nucleotide variant; *SQLE*, squalene epoxidase; SREBP, sterol regulatory element-binding protein; SS, squalene synthase; TAD, topologically associating domain; TSSs, transcription start sites.

Availability of Data and Materials

All data and datasets on which the conclusions of the manuscript depend are presented in the study or are provided as part of the **Supplementary Material**. The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Author Contributions

EI designed the research study. EI and AM performed the research. EI and AM analyzed the data. EI wrote the manuscript. Both authors contributed to editorial changes in the manuscript. Both authors read and approved the final manuscript. Both authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

Not applicable.

Acknowledgment

We are grateful to all peer reviewers for their opinions and valuable remarks.

Funding

This research was supported by publicly funded project No. FWNR-2026-0023 of the Federal Research Center ICG SB RAS.

Conflicts of Interest

The authors declare no conflicts of interest.

Declaration of AI and AI-Assisted Technologies in the Writing Process

During the preparation of this work, the authors used Google and Yandex translators. After using these tools, the paper was professionally edited. The authors reviewed the text obtained and take full responsibility for the content of the publication.

Supplementary Material

Supplementary Table 1. Databases and query terms that were used to search for publications containing data on SREBP-binding sites. **Supplementary Table 2**. SREBP-binding sites in the regulatory regions of genes from the gene network controlling cholesterol biosynthesis extracted from scientific publications (extended Table 1). **Supplementary Table 3**. Deletions affecting SREBP-binding sites. **Supplementary Table 4**. 171 SNVs colocalized with SREBP sites or the nearest three left or three right flanking nucleotides. **Supplementary Table 5**. The predicted effect of four single-nucleotide deletions and five top-ranked SNVs on the similarity of SREBP-binding sites to SREBP1 and SREBP2 motif models (matrices) from the HOCOMOCO database. **Supplementary Table 6**. 20 SNVs that have the highest ranks according to the PhyloP scores of affected nucleotides. Supplementary material associated with this article can be found, in the online version, at https://doi.org/10.31083/FBS51759.

References

[1] Schade DS, Shey L, Eaton RP. Cholesterol Review: A Metabolically Important Molecule. Endocrine Practice: Official Journal of the American College of Endocrinology and the American Association of Clinical Endocrinologists. 2020; 26: 1514–1523. https://doi.org/10.4158/EP-2020-0347

[2] Jesch ED, Carr TP. Food Ingredients That Inhibit Cholesterol Absorption. Preventive Nutrition and Food Science. 2017; 22: 67–80. https://doi.org/10.3746/pnf.2017.22.2.67

[3] Luo J, Yang H, Song BL. Mechanisms and regulation of cholesterol homeostasis. Nature Reviews. Molecular Cell Biology. 2020; 21: 225–245. https://doi.org/10.1038/s41580-019-0190-7

[4] Mitsche MA, McDonald JG, Hobbs HH, Cohen JC. Flux analysis of cholesterol biosynthesis in vivo reveals multiple tissue and cell-type specific pathways. ELife. 2015; 4: e07999. https://doi.org/10.7554/eLife.07999

[5] Genaro-Mattos TC, Anderson A, Allen LB, Korade Z, Mirnics K. Cholesterol Biosynthesis and Uptake in Developing Neurons. ACS Chemical Neuroscience. 2019; 10: 3671–3681. https://doi.org/10.1021/acschemneuro.9b00248

[6] Sharpe LJ, Brown AJ. Controlling cholesterol synthesis beyond 3-hydroxy-3-methylglutaryl-CoA reductase (HMGCR). The Journal of Biological Chemistry. 2013; 288: 18707–18715. https://doi.org/10.1074/jbc.R113.479808

[7] Slenter D, Weitz E, Hanspers K. Cholesterol metabolism with Bloch and Kandutsch-Russell pathways (WP4718). 2023. Available at: https://www.wikipathways.org/instance/WP4718 (Accessed: February 10, 2026).

[8] Waller DD, Park J, Tsantrizos YS. Inhibition of farnesyl pyrophosphate (FPP) and/or geranylgeranyl pyrophosphate (GGPP) biosynthesis and its implication in the treatment of cancers. Critical Reviews in Biochemistry and Molecular Biology. 2019; 54: 41–60. https://doi.org/10.1080/10409238.2019.1568964

[9] Gunasekaran M, Littel HR, Wells NM, Turner J, Campos G, Venigalla S, et al. Effects of HMG CoA reductase (HMGCR) deficiency on skeletal muscle development. The FEBS Journal. 2025; 292: 4854–4869. https://doi.org/10.1111/febs.17406

[10] Nagashima S, Yagyu H, Ohashi K, Tazoe F, Takahashi M, Ohshiro T, et al. Liver-specific deletion of 3-hydroxy-3-methylglutaryl coenzyme A reductase causes hepatic steatosis and death. Arteriosclerosis, Thrombosis, and Vascular Biology. 2012; 32: 1824–1831. https://doi.org/10.1161/ATVBAHA.111.240754

[11] Yogev Y, Shorer Z, Koifman A, Wormser O, Drabkin M, Halperin D, et al. Limb girdle muscular disease caused by

HMGCR mutation and statin myopathy treatable with mevalonolactone. Proceedings of the National Academy of Sciences of the United States of America. 2023; 120: e2217831120. https://doi.org/10.1073/pnas.2217831120

[12] Tiniakou E. Statin-Associated Autoimmune Myopathy: Current Perspectives. Therapeutics and Clinical Risk Management. 2020; 16: 483–492. https://doi.org/10.2147/TCRM.S197941

[13] Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature. 2010; 466: 707–713. https://doi.org/10.1038/nature09270

[14] Chen YC, Chen YDI, Li X, Post W, Herrington D, Polak JF, et al. The HMG-CoA reductase gene and lipid and lipoprotein levels: the multi-ethnic study of atherosclerosis. Lipids. 2009; 44: 733–743. https://doi.org/10.1007/s11745-009-3314-6

[15] Kolchanov NA, Ignatieva EV, Podkolodnaya OA, Likhoshvai VA, Matushkin YuG. Gene Networks. Vavilov Journal of Genetics and Breeding. 2013; 17: 833–850. (In Russian)

[16] Mikhailova AD, Lashin SA, Ivanisenko VA, Demenkov PS, Ignatieva EV. Reconstruction and computer analysis of the structural and functional organization of the gene network regulating cholesterol biosynthesis in humans and the evolutionary characteristics of the genes involved in the network. Vavilovskii Zhurnal Genetiki I Selektsii. 2024; 28: 864–873. https://doi.org/10.18699/vjgb-24-94

[17] Jiang T, Zhang G, Lou Z. Role of the Sterol Regulatory Element Binding Protein Pathway in Tumorigenesis. Frontiers in Oncology. 2020; 10: 1788. https://doi.org/10.3389/fonc.2020.01788

[18] Hua X, Wu J, Goldstein JL, Brown MS, Hobbs HH. Structure of the human gene encoding sterol regulatory element binding protein-1 (SREBF1) and localization of SREBF1 and SREBF2 to chromosomes 17p11.2 and 22q13. Genomics. 1995; 25: 667–673. https://doi.org/10.1016/0888-7543(95)80009-b

[19] Miserez AR, Cao G, Probst LC, Hobbs HH. Structure of the human gene encoding sterol regulatory element binding protein 2 (SREBF2). Genomics. 1997; 40: 31–40. https://doi.org/10.1006/geno.1996.4525

[20] Sato R, Inoue J, Kawabe Y, Kodama T, Takano T, Maeda M. Sterol-dependent transcriptional regulation of sterol regulatory element-binding protein-2. The Journal of Biological Chemistry. 1996; 271: 26461–26464. https://doi.org/10.1074/jbc.271.43.26461

[21] Raghow R, Yellaturu C, Deng X, Park EA, Elam MB. SREBPs: the crossroads of physiological and pathological lipid homeostasis. Trends in Endocrinology and Metabolism: TEM. 2008; 19: 65–73. https://doi.org/10.1016/j.tem.2007.10.009

[22] Smith JR, Osborne TF, Goldstein JL, Brown MS. Identification of nucleotides responsible for enhancer activity of sterol regulatory element in low density lipoprotein receptor gene. The Journal of Biological Chemistry. 1990; 265: 2306–2310.

[23] Kim JB, Spotts GD, Halvorsen YD, Shih HM, Ellenberger T, Towle HC, et al. Dual DNA binding specificity of ADD1/SREBP1 controlled by a single amino acid in the basic helix-loop-helix domain. Molecular and Cellular Biology. 1995; 15: 2582–2588. https://doi.org/10.1128/MCB.15.5.2582

[24] Vorontsov IE, Eliseeva IA, Zinkevich A, Nikonov M, Abramov S, Boytsov A, et al. HOCOMOCO in 2024: a rebuild of the curated collection of binding models for human and mouse transcription factors. Nucleic Acids Research. 2024; 52: D154–D163. https://doi.org/10.1093/nar/gkad1077

[25] Sakakura Y, Shimano H, Sone H, Takahashi A, Inoue N, Toyoshima H, et al. Sterol regulatory element-binding proteins induce an entire pathway of cholesterol synthesis. Biochemical and Biophysical Research Communications. 2001; 286: 176–183. https://doi.org/10.1006/bbrc.2001.5375

[26] Horton JD, Shah NA, Warrington JA, Anderson NN, Park SW, Brown MS, et al. Combined analysis of oligonucleotide microarray data from transgenic and knockout mice identifies direct SREBP target genes. Proceedings of the National Academy of Sciences of the United States of America. 2003; 100: 12027–12032. https://doi.org/10.1073/pnas.1534923100

[27] Reed BD, Charos AE, Szekely AM, Weissman SM, Snyder M. Genome-wide occupancy of SREBP1 and its partners NFY and SP1 reveals novel functional roles and combinatorial regulation of distinct classes of genes. PLoS Genetics. 2008; 4: e1000133. https://doi.org/10.1371/journal.pgen.1000133

[28] Hennekam RCM. Congenital brain anomalies in distal cholesterol biosynthesis defects. Journal of Inherited Metabolic Disease. 2005; 28: 385–392. https://doi.org/10.1007/s10545-005-7055-2

[29] Andersson HC. Disorders of post-squalene cholesterol biosynthesis leading to human dysmorphogenesis. Cellular and Molecular Biology (Noisy-le-Grand, France). 2002; 48: 173–177.

[30] Jira P. Cholesterol metabolism deficiency. Handbook of Clinical Neurology. 2013; 113: 1845–1850. https://doi.org/10.1016/B978-0-444-59565-2.00054-X

[31] Atzmony L, Khan HM, Lim YH, Paller AS, Levinsohn JL, Holland KE, et al. Second-Hit, Postzygotic PMVK and MVD Mutations in Linear Porokeratosis. JAMA Dermatology. 2019; 155: 548–555. https://doi.org/10.1001/jamadermatol.2019.0016

[32] Zhang Z, Li C, Wu F, Ma R, Luan J, Yang F, et al. Genomic variations of the mevalonate pathway in porokeratosis. ELife. 2015; 4: e06322. https://doi.org/10.7554/eLife.06322

[33] Brunetti-Pierri N, Corso G, Rossi M, Ferrari P, Balli F, Rivasi F, et al. Lathosterolosis, a novel multiple-malformation/mental retardation syndrome due to deficiency of 3beta-hydroxysteroid-delta5-desaturase. American Journal of Human Genetics. 2002; 71: 952–958. https://doi.org/10.1086/342668

[34] Schafer BL, Bishop RW, Kratunis VJ, Kalinowski SS, Mosley ST, Gibson KM, et al. Molecular cloning of human mevalonate kinase and identification of a missense mutation in the genetic disease mevalonic aciduria. The Journal of Biological Chemistry. 1992; 267: 13229–13238.

[35] Houten SM, Romeijn GJ, Koster J, Gray RG, Darbyshire P, Smit GP, et al. Identification and characterization of three novel missense mutations in mevalonate kinase cDNA causing mevalonic aciduria, a disorder of isoprene biosynthesis. Human Molecular Genetics. 1999; 8: 1523–1528. https://doi.org/10.1093/hmg/8.8.1523

[36] Waterham HR, Koster J, Romeijn GJ, Hennekam RC, Vreken P, Andersson HC, et al. Mutations in the 3beta-hydroxysterol Delta24-reductase gene cause desmosterolosis, an autosomal recessive disorder of cholesterol biosynthesis. American Journal of Human Genetics. 2001; 69: 685–694. https://doi.org/10.1086/323473

[37] Ignatieva EV, Matrosova EA. Disease-associated genetic variants in the regulatory regions of human genes: mechanisms of action on transcription and genomic resources for dissecting these mechanisms. Vavilovskii Zhurnal Genetiki I Selektsii. 2021; 25: 18–29. https://doi.org/10.18699/VJ21.003

[38] Antontseva EV, Degtyareva AO, Korbolina EE, Damarov IS, Merkulova TI. Human-genome single nucleotide polymorphisms affecting transcription factor binding and their role in pathogenesis. Vavilovskii Zhurnal Genetiki I Selektsii. 2023; 27: 662–675. https://doi.org/10.18699/VJGB-23-77

[39] Kolchanov NA, Ignatieva EV, Ananko EA, Podkolodnaya OA, Stepanenko IL, Merkulova TI, et al. Transcription Regulatory Regions Database (TRRD): its status in 2002. Nucleic Acids Research. 2002; 30: 312–317. https://doi.org/10.1093/nar/30.1.312

[40] Han H, Shim H, Shin D, Shim JE, Ko Y, Shin J, et al. TRRUST: a reference database of human transcriptional regulatory inter-

actions. Scientific Reports. 2015; 5: 11432. https://doi.org/10.1038/srep11432
[41] Kent WJ. BLAT—the BLAST-like alignment tool. Genome Research. 2002; 12: 656–664. https://doi.org/10.1101/gr.229202
[42] Kuhn RM, Haussler D, Kent WJ. The UCSC genome browser and associated tools. Briefings in Bioinformatics. 2013; 14: 144–161. https://doi.org/10.1093/bib/bbs038
[43] Ramani R, Krumholz K, Huang YF, Siepel A. PhastWeb: a web interface for evolutionary conservation scoring of multiple sequence alignments using phastCons and phyloP. Bioinformatics (Oxford, England). 2019; 35: 2320–2322. https://doi.org/10.1093/bioinformatics/bty966
[44] Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. Genome Research. 2010; 20: 110–121. https://doi.org/10.1101/gr.097857.109
[45] Taher L, Narlikar L, Ovcharenko I. Identification and computational analysis of gene regulatory elements. Cold Spring Harbor Protocols. 2015; 2015: pdb.top083642. https://doi.org/10.1101/pdb.top083642
[46] Kulakovskiy IV, Vorontsov IE, Yevshin IS, Sharipov RN, Fedorova AD, Rumynskiy EI, et al. HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. Nucleic Acids Research. 2018; 46: D252–D259. https://doi.org/10.1093/nar/gkx1106
[47] Shimano H, Sato R. SREBP-regulated lipid metabolism: convergent physiology - divergent pathophysiology. Nature Reviews. Endocrinology. 2017; 13: 710–730. https://doi.org/10.1038/nrendo.2017.91
[48] Tansey TR, Shechter I. Squalene synthase: structure and regulation. Progress in Nucleic Acid Research and Molecular Biology. 2001; 65: 157–195. https://doi.org/10.1016/s0079-6603(00)65005-5
[49] DeBose-Boyd RA, Ye J. SREBPs in Lipid Metabolism, Insulin Signaling, and Beyond. Trends in Biochemical Sciences. 2018; 43: 358–368. https://doi.org/10.1016/j.tibs.2018.01.005
[50] Párraga A, Bellsolell L, Ferré-D'Amaré AR, Burley SK. Co-crystal structure of sterol regulatory element binding protein 1a at 2.3 A resolution. Structure (London, England: 1993). 1998; 6: 661–672. https://doi.org/10.1016/s0969-2126(98)00067-7
[51] Seo YK, Jeon TI, Chong HK, Biesinger J, Xie X, Osborne TF. Genome-wide localization of SREBP-2 in hepatic chromatin predicts a role in autophagy. Cell Metabolism. 2011; 13: 367–375. https://doi.org/10.1016/j.cmet.2011.03.005
[52] Rohs R, Jin X, West SM, Joshi R, Honig B, Mann RS. Origins of specificity in protein-DNA recognition. Annual Review of Biochemistry. 2010; 79: 233–269. https://doi.org/10.1146/annurev-biochem-060408-091030
[53] Oshchepkov DY, Vityaev EE, Grigorovich DA, Ignatieva EV, Khlebodarova TM. SITECON: a tool for detecting conservative conformational and physicochemical properties in transcription factor binding site alignments and for site recognition. Nucleic Acids Research. 2004; 32: W208–W212. https://doi.org/10.1093/nar/gkh474
[54] Ignatieva EV, Merkulova TI, Oshchepkov DY, Klimova NV, Vasiliev GV, Turnaev II, et al. Prediction of New SREBP Binding Sites in the Promoter Regions of Vertebrate Genes Using a Combination of a Bioinformatic and an Experimental Approach. Information Bulletin of VOGIS. 2009; 13: 37–45. (In Russian)
[55] Thompson JA, Congdon CB. An exploration into improving DNA motif inference by looking for highly conserved core regions. In 2013 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB) (pp. 60–67). Singapore. 2013. https://doi.org/10.1109/CIBCB.2013.6595389
[56] 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 491: 56–65. https://doi.org/10.1038/nature11632
[57] Peña-Martínez EG, Rodríguez-Martínez JA. Decoding Non-coding Variants: Recent Approaches to Studying Their Role in Gene Regulation and Human Diseases. Frontiers in Bioscience (Scholar Edition). 2024; 16: 4. https://doi.org/10.31083/j.fbs1601004
[58] Degtyareva A, Antontseva E, Evseenko A, Orishchenko K, Merkulova T. The Single Nucleotide Substitution T → A rs2072580 Damages the CREB1 Binding Site in the Bidirectional SART3/ISCU Promoter. Genes. 2025; 16: 713. https://doi.org/10.3390/genes16060713
[59] Brown CD, Mangravite LM, Engelhardt BE. Integrative modeling of eQTLs and cis-regulatory elements suggests mechanisms underlying cell type specificity of eQTLs. PLoS Genetics. 2013; 9: e1003649. https://doi.org/10.1371/journal.pgen.1003649
[60] Vohra M, Sharma AR, Prabhu B N, Rai PS. SNPs in Sites for DNA Methylation, Transcription Factor Binding, and miRNA Targets Leading to Allele-Specific Gene Expression and Contributing to Complex Disease Risk: A Systematic Review. Public Health Genomics. 2020; 23: 155–170. https://doi.org/10.1159/000510253
[61] Korbolina EE, Bryzgalov LO, Ustrokhanova DZ, Postovalov SN, Poverin DV, Damarov IS, et al. A Panel of rSNPs Demonstrating Allelic Asymmetry in Both ChIP-seq and RNA-seq Data and the Search for Their Phenotypic Outcomes through Analysis of DEGs. International Journal of Molecular Sciences. 2021; 22: 7240. https://doi.org/10.3390/ijms22147240
[62] Mudappathi R, Patton T, Chen H, Yang P, Sun Z, Wang P, et al. reg-eQTL: Integrating transcription factor effects to unveil regulatory variants. American Journal of Human Genetics. 2025; 112: 659–674. https://doi.org/10.1016/j.ajhg.2025.01.015
[63] Degtyareva AO, Antontseva EV, Merkulova TI. Regulatory SNPs: Altered Transcription Factor Binding Sites Implicated in Complex Traits and Diseases. International Journal of Molecular Sciences. 2021; 22: 6454. https://doi.org/10.3390/ijms22126454
[64] Coman D, Vissers LELM, Riley LG, Kwint MP, Hauck R, Koster J, et al. Squalene Synthase Deficiency: Clinical, Biochemical, and Molecular Characterization of a Defect in Cholesterol Biosynthesis. American Journal of Human Genetics. 2018; 103: 125–130. https://doi.org/10.1016/j.ajhg.2018.05.004
[65] Guan G, Jiang G, Koch RL, Shechter I. Molecular Cloning and Functional Analysis of the Promoter of the Human Squalene Synthase Gene. Journal of Biological Chemistry. 1995; 270: 21958–21965. https://doi.org/10.1074/jbc.270.37.21958
[66] Tadaka S, Katsuoka F, Ueki M, Kojima K, Makino S, Saito S, et al. 3.5KJPNv2: an allele frequency panel of 3552 Japanese individuals including the X chromosome. Human Genome Variation. 2019; 6: 28. https://doi.org/10.1038/s41439-019-0059-5
[67] Jeon S, Bhak Y, Choi Y, Jeon Y, Kim S, Jang J, et al. Korean Genome Project: 1094 Korean personal genomes with clinical information. Science Advances. 2020; 6: eaaz7835. https://doi.org/10.1126/sciadv.aaz7835
[68] Howe V, Sharpe LJ, Prabhu AV, Brown AJ. New insights into cellular cholesterol acquisition: promoter analysis of human HMGCR and SQLE, two key control enzymes in cholesterol synthesis. Biochimica et Biophysica Acta. Molecular and Cell Biology of Lipids. 2017; 1862: 647–657. https://doi.org/10.1016/j.bbalip.2017.03.009