Whole-exome sequencing reveals migraine associated novel functional variants

Lubna Ibrahim Al Asoom
Imam Abdulrahman Bin Faisal University
Johra Khan (j.khan@mu.edu.sa)
Majmaah University College of Applied Medical Sciences https://orcid.org/0000-0002-0044-4758
Ahmad Al Sunni
Imam Abdulrahman Bin Faisal University
Nazish Rafique
Imam Abdulrahman Bin Faisal University
Rabia Latif
Imam Abdulrahman Bin Faisal University
Majed M. Alabdali
Imam Abdulrahman Bin Faisal University
Azhar Alhariri
Imam Abdulrahman Bin Faisal University
Majed Aloqaily
Imam Abdulrahman Bin Faisal University
Sayed AbdulAzeez
Imam Abdulrahman Bin Faisal University
J. Francis Borgio
Imam Abdulrahman Bin Faisal University

Research Article

Keywords: Exome, Migraine, Whole Exome Sequencing, RETNLB, Saudi Arabia, Next Generation Sequencing, Risk Variants

Posted Date: September 16th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-885676/v1

Abstract

Background

Migraine, as the 7th most disabling neurological multi-symptomatic disease condition and 26.9% prevalence in Saudi females lacks studies on SNPs for their relation with migraine aura.

Methods

This study was conducted on 40 Arab ancestry young female subjects, among whom 50% cases with migraine and remaining controls were used to identify the migraine associated novels genes and risk variants. After quality controls, 3365343 missense, frameshift, missense splice region variants and insertion-deletion (indels) polymorphisms were tested for association to migraine.

Results

Seventeen significant (*p* value 9.091×10^{-05}) functional variants in 12 genes (*RETNLB, SCAI, ADH4, ESPL1, CPT2, FLG, PPP4R1, SERPINB5, ZNF66, ETAA1, EXO1* and *CPA6*) were migraine risk associated including a stop gained frameshift (-13-14*SX) variant in the gene *RETNLB* (rs5851607; *p* value 3.446×10^{-06}). Gene analysis revealed that half of the significant novel migraine risk genes expressed in temporal lobe (*p*-value 0.0058) of the cerebral cortex.

Conclusions

This first study in female (22.10 ± 3.63 years) migrainers is the first one exploring migraine risk variants in Arab ancestry.

Background

Migraine is a multifactorial chronic neurological disorder with a global prevalence between 14 to 21.7%. It creates a burden on the economy as well as on the normal lives of individuals [1]. Migraine is the 7th most disabling disease around the world as it causes 2.9% Years of life lost to disability (YLD) [2]. Out of 4 million adults, 85% includes female adults suffering from chronic headaches [3]. In Saudi Arabia, the prevalence of migraines in females is around 26.9% [4]. Recent studies revealed the high estrogen levels on mediating non-menstrual related migraine among young Saudi females and serum ApoE has was reported as excellent diagnostic marker for Saudi females with migraine inictal or interictal phase [5]. Recently, a cross-sectional population-based study in Saudi Arabia reveals the prevalence of migraine is higher than the global averages [6].

Migraine is classified mainly into two forms: migraine with aura (MA) and migraine without aura (MO). It can also be classified into chronic and episodic migraine. Hemiplegic migraine is another type of MA and is a severe and rare condition that affects one side of the body and causes temporary numbness [7]. The old theory on the cause of migraine considered it as a vascular disease; However, many recent researches confirmed the involvement of multiple mechanisms related to the brain structures and processes [8]. Furthermore, accumulated evidence demonstrates a genetic background for the transmission of the disease. Many epidemiological studies showed

the passage of migraine in families with a concordance rate of 1 to 2 folds in monozygotic twins in comparison to dizygotic twins [9]. Different approaches were used to determine the genetic factors, and the DNA variants that might be responsible for the migraine. Some studies focused on candidate gene associated studies (CGAS) to identify specific genetic markers associated with migraine but the negative result of these studies and the development of cost-effective techniques resulted in genome-wide association studies (GWAS) [10]. In these studies, hundreds of thousands of single nucleotide polymorphisms (SNPs) were screened for their association with migraine. Studies from headache clinics of the Netherland, Germany, and Finland on 10,747 population leads to the identification of rs835740, a single significant SNP [11]. A similar recent study identified 38 different SNPs and seven loci *TRPM8, LRP1, FHL5, TSPAN2, ASTN2*, near *FGF6*, and *PHACTR1* but none of them were found related to migraine without aura [12]. Focused studies on SNPs especially in females are needed, as they are the most affected patients of migraine. Identification of these migraine associated SNP needs collaboration from different areas of the world in order to collect sufficient information from different human races and draw the full picture of the genetic predisposition of migraine. Therefore, this study was designed to identify the associated gene variants and single nucleotide polymorphisms with migraine in young Saudi females using whole exome sequencing analysis.

Methods

The study protocol was approved by the Institution Review Board of Imam Abdulrahman Bin Faisal University (IRB number: IRB-2021-01-250) and was conducted in accordance with the Declaration of Helsinki. It is a cross-sectional case-control study conducted in the female campus of College of Medicine and the Institute for Research and Medical Consultations (IRMC) of Imam Abdulrahman bin Faisal University (IAU), Dammam, Saudi Arabia. The study involved 40 participants: 20 controls (healthy subjects) and 20 cases (migraineurs). Subjects were recruited by convenience sampling and were all Saudi female college students with age range: 18–30. For the cases to be included in the study, they had to be diagnosed with migraine by a neurologist and satisfy the criteria of ICHD. The International Classification of Headache Disorders 3rd edition was considered as diagnosis reference for selecting patients with headache. Controls were healthy female subjects of the same age group and with no complaints of headache.

All participants filled a written informed consent for their enrolment in the study, then they were interviewed and asked to fill an electronic data sheet (**Supplementary Material 1**). The data sheet includes: the demographic characteristics (age, marital status, college level, height, weight, BMI). In addition, it includes specific questions related to migraine such as: frequency of attacks/month, severity of attack (using visual scale 1–10), associated symptoms, presence of triggers (stress, lack of sleep, missed meal or fasting, physical activity, noise, smell, strong lights, fluctuation of weather or temperature, food, relation to menstrual cycle, type of migraine (with aura, without aura), family history of migraine, past history (other chronic diseases), use of medications (for migraine, for other diseases).

Whole exome sequencing of migraine and control samples and statistical data analysis

Blood samples were obtained from the study participants in EDTA-vacutainers. Deoxyribonucleic acid was extracted using QIAamp DNA Blood Mini Kit (Qiagen, Germany) and DNA purity was checked using nanodrop and concentration was determined using qubit fluorometer. DNA integrity was tested using agarose gel electrophoresis. Paired end whole exome sequencing was done for all the samples and subjected for quality screening. Sample depth mean ≥ 7.5 and sample variant call rate ≥ 0.5 and sample genotype quality mean ≥ 28

were considered for good quality sample. The conditions like phred score quality ≥ 30, raw read depth ≥ 10 and mapping quality ≥ 30 were applied for filtering the good quality variants. The Hail standard (python package) was followed for the entire genome-wide association study (GWAS) pipeline analysis. Similar python package was used for the quality check of the samples, variants and genotypes. Manhattan plot and QQ plot of the association of SNPs with migraine as statistical significance in terms of *p*-values on a genomic scale was constructed. Gene level linkage disequilibrium analysis of the SNPs in the significant gene was done using Haploview 4.2. Data were analysed using statistical package for the social sciences (SPSS) software version 21.

The most significant top 10 genes analysed using the gene functional classification tool DAVID to identify the significance in the site of expression. The functional annotation of top 50 highly associated genes (with lowest *p*-value < 0.00023), was performed using Uniprot database. Further the GO and Pathway enrichment was carried using enrichR server and pathway involvement of the genes were done by KEGG search and color pathway server. All the significant markers identified from the GWAS and genes were selected for the expression profile analysis in the brain and related tissues using DAVID. Brain and related tissue expressed genes were separated and analysed for KEGG pathway enrichment.

Results

Study Population

Study participants (Table 1) were drawn from Arab ancestries. The demographic characteristics of the 40 selected Saudi Arabian subjects including 20 migraine patients and controls (n = 20) are presented in Table 1. The clinical characteristics and the frequency of the precipitating factors of headache attack in the migraineurs of the study is presented in Table 2.

Table 1
Demographic characteristics of the migraineurs and the controls.

	Control n = 20	*Migraineurs n = 20*	*p value**
Age (years)	21.86 ± 1.75	22.10 ± 3.63	0.818
Body weight (Kg)	56.21 ± 14.02	63.00 ± 12.58	0.155
BMI	21.48 ± 5.39	24.66 ± 5.18	0.095
*Student t-test			

Table 2
Clinical characteristics of the migraineurs of the study and the frequency of the precipitating factors:

Variable	Description	
Type of migraine	With aura 11 (57.9%), Without aura 9 (42.1%)	
Family history	Yes 11 (55%), No 9 (45%)	
Use of medications	No medication 7 (35%), pain killer 10 (50%), prophylaxis 3 (15%)	
Number of attacks/months	Ranges from 2-28attacks/month with 2 attacks/month is the most common	
Duration of the attack	Ranges from 4 hours to > 72 hours with the most common is 12–24 hours	
Severity of the attack (Visual scale 1–10)	Ranges from 7–10	
	Precipitating factors	
	Yes	**No**
Sleep disturbances	17 (85%)	3 (15%)
Stress	16 (80%)	4 (20%)
Bright light	18 (90%)	2 (10%)
Excessive noise	15 (75%)	5 (25%)
Strong smells	12 (60%)	8 (40%)
Weather changes	8 (40%)	12 (60%)
Skipped meal	10 (50%)	10 (50%)
Physical exertion	5 (25%)	15 (75%)
Certain type of food	5 (25%)	15 (75%)
Coffee	2 (10%)	18 (90%)

Single-variant Analysis

After quality controls for the variants obtained in the whole exome sequencing, 3365343 variants were satisfied for further GWAS analyses. Our study highlights the added influence of considering the functional variants like missense variants, frameshift variant and missense splice region variant in the analysis: set-level p-values < 0.00001 from GWAS analysis (Table 3) to prioritize the top migraine associated variants in Arab ancestry. Entire list of migraine associated (*p* value < 0.00001) variants identified through exome sequencing are shown in **Supplementary Table 1**. Seventeen variants were found to be the most significant (9.091×10^{-05}) functional variants distributed among 12 genes (*RETNLB, SCAI, ADH4, ESPL1, CPT2, FLG, PPP4R1, SERPINB5, ZNF66, ETAA1, EXO1* and *CPA6*) (Table 3; Fig. 1). The stop gained frameshift (-13-14*SX) variant in the gene *RETNLB* is the most significant functional variant (rs5851607; *p* value = 3.446×10^{-06}).

Table 3

List of migraine associated (*p* value < 0.00001) exonic functional variants identified through exome sequencing.

	Locus. contig	Variation ID	Gene SYMBOL	MAF	Alleles	Amino acids	Codons	*p*-value
1	chr3	rs5851607*	*RETNLB*	0.18	['G', 'GGGGGATTA']	-13-14*SX	-/TAATCCCC	3.446×10^{-06}
2	chr9	rs589292	*SCAI*	0.2	['C', 'T']	A37T	Gct/Act	1.169×10^{-05}
3	chr4	rs1126671	*ADH4*	0.26	['T', 'C']	I309V	Att/Gtt	1.575×10^{-05}
4	chr4	rs1126673$^{\$}$	*ADH4*	0.26	['C', 'T']	V393I	Gtc/Atc	1.575×10^{-05}
5	chr12	rs6580942	*ESPL1*	0.3	['C', 'A']	A25D	gCc/gAc	3.413×10^{-05}
6	chr1	rs1799821	*CPT2*	0.28	['G', 'A']	V368I	Gtc/Atc	3.585×10^{-05}
7	chr1	rs3126075	*FLG*	0.23	['G', 'C']	T3579R	aCg/aGg	4.266×10^{-05}
8	chr18	rs329003	*PPP4R1*	0.28	['T', 'C']	I381V	Ata/Gta	4.500×10^{-05}
9	chr18	rs2289520	*SERPINB5*	0.14	['G', 'C']	V187L	Gtc/Ctc	4.556×10^{-05}
10	chr12	rs56358776	*ESPL1*	0.26	['G', 'A']	R1561Q	cGg/cAg	5.923×10^{-05}
11	chr19	rs10413187	*ZNF66*	0.14	['C', 'A']	Q66K	Cag/Aag	7.328×10^{-05}
12	chr19	rs432839	*ZNF66*	0.22	['G', 'T']	C173F	tGc/tTc	7.320×10^{-05}
13	chr19	rs383038	*ZNF66*	0.14	['T', 'C']	F188L	Ttt/Ctt	7.328×10^{-05}
14	chr19	rs370551	*ZNF66*	0.14	['A', 'G']	T420A	Act/Gct	7.328×10^{-05}
15	chr2	rs61740794	*ETAA1*	0.4	['G', 'A']	E673K	Gaa/Aaa	8.508×10^{-05}

* Stop gained frameshift variant; $^{\$}$ missense splice region variant. Significant (p value < 0.00001) functional and other variants identified through exome sequencing are listed in the **Table S1**. MAF: Minor allele frequency.

	Locus. contig	Variation ID	Gene SYMBOL	MAF	Alleles	Amino acids	Codons	*p*-value
16	chr1	rs735943	*EXO1*	0.42	['A', 'G']	H354R	cAt/cGt	8.978×10^{-05}
17	chr8	rs17343819[$]	*CPA6*	0.22	['T', 'C']	N249S	aAt/aGt	9.091×10^{-05}

* Stop gained frameshift variant; [$] missense splice region variant. Significant (p value < 0.00001) functional and other variants identified through exome sequencing are listed in the **Table S1**. MAF: Minor allele frequency.

Genes Analysis

The most significant top 12 genes analysed using the gene functional classification tool, DAVID to identify the significance in the site of expression. The analysis revealed that six out of 12 genes (Table 3) significantly expressed in temporal lobe (*p*-value = 0.00582) (Fig. 2). The functional variants with *p*-value between 9.091×10^{-05} to 0.05 on the genes identified as expressed in temporal lobe revealed the significance of *FLG* gene with 37 functional variants (Table 4). The significant SNPs observed in the *FLG* gene and their amino acid position are presented in Fig. 3. The gene level linkage disequilibrium analysis (Haploview 4.2) of the 53 SNPs in the *FLG* gene reveals significant association for all these SNPs (Chi Square = 8.556; *p*-value = 0.0034) (Fig. 3). The most significant three markers haplotype rs3126075G, rs7532285T and rs7540123G (Chi Square = 7.64; *p*-value = 0.0057) appear to associate significantly with migraine; while the opposite alleles rs3126075C, rs7532285C and rs7540123C (Chi Square = 3.81; *p*-value = 0.0407) is protective type of haplotype. Presence of variants at *FLG* were confirmed using Sanger sequencing with the designed primers (Forward primer: FLGF: 5′ CCTCTACCAGGTGAGCACTCATGAACAGTCTG 3′ and Reverse primer: FLGR: 5′ TCTCTGACTGCAGATGAAGCTTGTCCGTGCC 3′). Sanger sequencing revealed additional variants in the gene and need to check their association with migraine (Fig. 4).

Table 4
List of significant migraine associated missense variants of *FLG* gene in the chromosome 1.

	SNP	alleles	MAF	Amino acids	Protein position	Codons	*p*-value
1	rs3126075	['G', 'C']	0.23	T/R	3579	aCg/aGg	4.27×10^{-05}
2	rs7532285	['T', 'C']	0.05	Q/R	3568	cAg/cGg	0.000148
3	rs7540123	['G', 'C']	0.05	Q/E	3568	Cag/Gag	0.000148
4	rs2065955	['C', 'G']	0.3	G/A	3436	gGa/gCa	0.000592
5	rs3126079	['G', 'T']	0.3	H/Q	1961	caC/caA	0.000592
6	rs58001094	['G', 'C']	0.3	A/G	1167	gCa/gGa	0.000592
7	rs11582087	['T', 'G']	0.04	S/R	2836	Agt/Cgt	0.00092
8	rs71625202	['C', 'G']	0.08	S/T	2366	aGt/aCt	0.00092
9	rs139476473	['C', 'T']	0.04	D/N	2339	Gac/Aac	0.00092
10	rs2065957	['A', 'C']	0.17	V/G	3179	gTg/gGg	0.001166
11	rs12083389	['C', 'G']	0.15	E/D	3593	gaG/gaC	0.002327
12	rs3126072	['C', 'T']	0.22	G/R	2545	Gga/Aga	0.003517
13	rs3126074	['G', 'C']	0.22	H/Q	2507	caC/caG	0.003517
14	rs2011331	['T', 'C']	0.22	T/A	454	Aca/Gca	0.003517
15	rs66954353	['T', 'G']	0.13	K/Q	2192	Aaa/Caa	0.003521
16	rs2184953	['A', 'G']	0.32	Y/H	2194	Tat/Cat	0.004922
17	rs140376327	['G', 'A']	0.04	R/W	2430	Cgg/Tgg	0.010229
18	rs12135040	['C', 'G']	0.04	G/R	1936	Ggg/Cgg	0.010229
19	rs138721961	['C', 'T']	0.04	R/H	402	cGc/cAc	0.010229
20	rs78179835	['C', 'G']	0.08	E/D	2297	gaG/gaC	0.014364
21	rs113544881	['A', 'T']	0.07	L/H	1943	cTt/cAt	0.016921
22	rs74129452	['T', 'G']	0.18	Q/H	2154	caA/caC	0.018862
23	rs7512553	['A', 'G']	0.18	Y/H	2119	Tat/Cat	0.018862
24	rs7522925	['G', 'A']	0.18	A/V	2108	gCg/gTg	0.018862
25	rs7512857	['A', 'C']	0.18	S/A	2020	Tca/Gca	0.018862
26	rs12407807	['C', 'T']	0.16	R/H	1684	cGc/cAc	0.023383
27	rs75235053	['C', 'G']	0.06	S/T	3662	aGt/aCt	0.026472
28	rs199888588	['A', 'G']	0.1	W/R	962	Tgg/Cgg	0.027008

MAF: Minor allele frequency.

	SNP	alleles	MAF	Amino acids	Protein position	Codons	*p*-value
29	rs74129455	['T', 'G']	0.1	K/Q	2064	Aaa/Caa	0.031164
30	rs149817134	['G', 'T']	0.04	H/N	1880	Cac/Aac	0.03179
31	rs55650366	['A', 'G']	0.14	L/S	2481	tTg/tCg	0.035772
32	rs71625200	['T', 'C']	0.14	K/E	2444	Aag/Gag	0.035772
33	rs71625201	['C', 'G']	0.14	E/Q	2398	Gag/Cag	0.035772
34	rs11581433	['T', 'C']	0.14	R/G	1376	Aga/Gga	0.035772
35	rs74129461	['C', 'T']	0.14	E/K	755	Gaa/Aaa	0.035772
36	rs11584340	['G', 'A']	0.14	P/S	478	Cct/Tct	0.035772
37	rs41267154	['C', 'A']	0.14	G/V	332	gGc/gTc	0.035772

MAF: Minor allele frequency.

Pathways Analysis

Gene ontology and Pathway enrichment of the top 50 genes revealed that the migraine subjects are moderately significant for the organic hydroxy compound catabolic process (GO:1901616; *p*-value = 9.12893E-05; adjusted *p*-value = 0.026108731) and quinone metabolic process (GO:1901661; *p*-value = 0.000272123; adjusted *p*-value = 0.038913602) (Table 5; **Supplementary Table 2**). All the significant markers having genes were checked for the expression profile in the in brain related tissues using DAVID. Total of 6305 genes were present in for DAVID among 11958 genes presented for expression analysis using DAVID. Total of 1349 genes were separated based on the expression in brain and related tissues were analysed for KEGG pathway enrichment, the results revealed that 34 genes were found to be significantly (Term *p*-value = 8.071×10^{-12}; adjusted *p*-value = 2.364×10^{-09}) associated with systemic lupus erythematosus (Table 6; **Supplementary Table 3; Figure S3).** Furthermore, the pathways like focal adhesion, ecm-receptor interaction, human papillomavirus infection, alcoholism, pathways in cancer, pi3k-akt signaling pathway, cholesterol metabolism are significantly associated with migraine in the Saudis (adjusted *p*-value $\leq 2.192\times10^{-05}$) (Table 5).

Table 5
KEGG pathway enrichment from the top 50 genes associated in the GWAS analysis.

Term	*p*-value	Number of Genes from top 50 list	Total Genes involved	Genes from top 50 list
p53 signaling pathway	0.000776571	3*	72	TNFRSF10B; SERPINB5; ATR
Fatty acid degradation	0.005418766	2[$]	44	ADH4;CPT2
Metabolism of xenobiotics by cytochrome P450	0.014752518	2	74	ADH4;AKR1C1
Cell cycle	0.038497072	2	124	ESPL1;ATR
Mannose type O-glycan biosynthesis	0.055975781	1	23	FKRP
Mismatch repair	0.055975781	1	23	EXO1
Tyrosine metabolism	0.086243722	1	36	ADH4
Human T-cell leukemia virus 1 infection	0.10395435	2	219	ESPL1;ATR
Vasopressin-regulated water reabsorption	0.104395242	1	44	AQP4
Fanconi anemia pathway	0.126588324	1	54	ATR
Steroid hormone biosynthesis	0.139644523	1	60	AKR1C1
Cytosolic DNA-sensing pathway	0.146100703	1	63	CCL4L2
Retinol metabolism	0.154635136	1	67	ADH4
Glycolysis / Gluconeogenesis	0.156755649	1	68	ADH4

* The p53 signaling pathway with the significant genes is presented in the **Figure S1**. * The Fatty acid degradation with the significant genes is presented in the **Figure S2**.

Table 6
KEGG pathway enrichment from the 1349 genes based on the expression in brain related tissues associated in the GWAS analysis.

Term	*p*-value	Adjusted *p*-value	Genes from 1349 list*	Total Genes
Systemic lupus erythematosus	8.071×10^{-12}	2.364×10^{-09}	34	133
Focal adhesion	2.665×10^{-11}	3.905×10^{-09}	42	199
ECM-receptor interaction	4.693×10^{-10}	4.583×10^{-08}	24	82
Human papillomavirus infection	3.167×10^{-09}	2.319×10^{-07}	53	330
Alcoholism	1.142×10^{-08}	6.693×10^{-07}	35	180
Pathways in cancer	2.191×10^{-08}	1.069×10^{-06}	71	530
PI3K-Akt signaling pathway	9.237×10^{-08}	3.866×10^{-06}	52	354
Cholesterol metabolism	5.986×10^{-07}	2.192×10^{-05}	15	50

* The list significant genes are presented in the **Supplementary Table 3**.

Discussion

Genetics plays a significant part in migraine in additions to other factors [13]. However, migraine genetic predisposition does not follow a direct Mendelian pattern. The common form of migraine is most probably polygenic and involves multiple variants at several genetic loci that possibly interact with multiple environmental factors. Genome-wide association studies (GWAS), are the most successful method to identify the genes involved in a disease. In this methodology, cohorts of migraine cases and controls are explored for any differences in allele frequencies of single nucleotide polymorphisms (SNPs) to identify genetic risk factors. There is no single genetic variant that can explain migraine heterogenecity across populations. We performed the first GWAS of migraine in Arab ancestry. Through exome sequencing we identified entire list of migraine associated (p value < 0.00001) variants and prioritised seventeen as the most significant (9.091×10^{-05}) functional variants distributed among 12 genes (*RETNLB, SCAI, ADH4, ESPL1, CPT2, FLG, PPP4R1, SERPINB5, ZNF66, ETAA1, EXO1* and *CPA6*) in the Saudi females suffering from migraine. All of these were novel and have not been documented in earlier studies involving other populations such as Europeans [14] and Chinese [15].

The stop gained frameshift variant in the gene *RETNLB* is the most significant functional variant, rs5851607, this gene encodes a bactericidal protein, Resistin-like molecule β (RELMβ) which is released from colonic cells to destroy Gram-negative bacteria (Propheter et al., 2017). Migraine may be associated with diseases such as irritable bowel syndrome (IBS), inflammatory bowel syndrome, and celiac disease [16]. *ADH4* gene encodes alcohol dehydrogenase enzyme and variations in this gene are associated with alcohol dependence [17]. *CPA6* gene encodes Carboxypeptidase A6 enzyme and its mutations can predispose to various types of epilepsy [18]. As shown in Fig. 2 **&** Table 3, a total of 36 functional variants were found to be significant in the gene, *FLG*. This gene encodes a protein called profilaggrin present in the epidermis of the skin. This protein is important for skin's barrier function. Functional variations in this gene can cause sensitization or atopic dermatitis [19] and might possibly be the underlying mechanism of Migraine-associated allodynia [20]. *ETAA1* gene that encodes a protein Ewing tumor-

associated antigen 1. This protein functions as a DNA replication stress response protein [21]. *CPT2* gene encodes carnitine palmitoyl transferase 2 enzyme which is essential for fatty acid oxidation. *ESPL1* gene codes a protease Separase/separin which causes separation of sister chromatids in mitosis [22]. *SERPINB5* gene encodes a protein Maspin, a tumour suppressor, that binds directly to extracellular matrix components and inhibits tumour-induced angiogenesis, invasion and metastatic spread [23]. SCAI encodes a protein that suppresses cancer cell invasion [24].

The morphological changes in the temporal lobe was reported to be associated with migraine [24]. Recently, reduction of gray matter volume in the temporal lobe was observed in migraine patients [25]. Gene analysis revealed that 6 of the significant 12 novel migraine risk genes expressed in temporal lobe of the cerebral cortex. Present study adds a molecular insight into the observations on the temporal lobe and migraine associated genes [26] and opens new avenues for migraine research. Current study will help in power calculations in future and will provide potential loci to look for in replication studies. This may facilitate thorough understanding of migraine pathophysiology and its underlying molecular mechanism, and open avenues for more precise diagnosis and therapeutic strategies targeting migraine patients of Arab ancestry. The study may also help in polygenic risk scoring, of the patients.

Conclusion

Our study is the first one exploring migraine genetic variations in Arab ancestry. Seventeen significant functional variants including a stop gained in 12 genes are the migraine risk variants in Arab ancestry. Half of the significant novel migraine risk genes expressed in temporal lobe of brain. The added migraine risk associated genes on the temporal lobe opens new opportunities for migraine pathophysiology and genetic research.

Abbreviations

MA
Migraine with Aura
MO
Migraine without Aura
YLD
Years of life lost to disability
CGAS
Candidate Gene Associated Studies
GWAS
Genome-Wide Association Studies
SNPs
Single Nucleotide Polymorphisms
IRMC
Institute for Research and Medical Consultations
IAU
Imam Abdulrahman bin Faisal University
ICHD
International Classification of Headache Disorders

RELMβ
Resistin-like moleculeβ
IBS
Irritable Bowel Syndrome
ADH4
*A*lcohol Dehydrogenase Enzyme
CPA6
Carboxypeptidase A6
ETAA1
Ewing Tumor-Associated Antigen 1
SCAI
Suppresses Cancer Cell Invasion

Declarations

Acknowledgments

The authors extend their appreciation to the Deputyship for Research & Innovation, Ministry of Education in Saudi Arabia for funding this research work through the project number (IFP-2020-46). The authors thank the Dean, Institute for Research and Medical Consultations (IRMC), Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia, for her continuous support and encouragement. The authors thank Mr.i87ytfdszsdty900ouQERTYUIOP

Ranilo, M. Tumbaga, and Mr. Horace T. Pacifico for their assistance. #][PIUTREWQ dfgnm,l;lkujhtrecvbn;.'

Authors' contributions

LAA, JK, SAA and JFB conceived, designed the research and analyzed the experiments. LAA, JK, AA, MAy, SAA, and JFB performed and analyzed the experiments. SAA, and JFB performed the whole genome analysis. LAA, AAS, NR, RL, and MAi performed clinical analysis. LAA, JK, and JFB wrote the paper with the contributions of AAS, NR, RL, MAi, AA, MAy, and SAA. LAA, JK, AAS, NR, RL, MAi, AA, MAy, SAA, and JFB reviewed and approved the final manuscript.

Funding

The Deputyship for Research & Innovation, Ministry of Education in Saudi Arabia funded this research work through the project number (IFP-2020-46).

Data Availability Statement and Data Citation

All data will be available on reasonable request from the corresponding author.

Ethics statement

The study was conducted in accordance with the Declaration of Helsinki and was approved by the Institutional Review Board (IRB) at Imam Abdulrahman Bin Faisal University. IRB approval number: IRB-2021-01-250. Written informed consent was obtained from each subject.

Competing interests

The authors declare that they have no competing interests.

References

1. Woldeamanuel YW, Cowan RP (2017) Migraine affects 1 in 10 people worldwide featuring recent rise: a systematic review and meta-analysis of community-based studies involving 6 million participants. Journal of the neurological sciences 372:307–315
2. Khan J, Al Asoom LI, Al Sunni A et al (2021) Genetics, pathophysiology, diagnosis, treatment, management, and prevention of migraine. Biomed Pharmacother 139:111557
3. Pietrobon D (2007) Familial hemiplegic migraine. Neurotherapeutics 4:274–284
4. Vos T, Flaxman AD, Naghavi M et al (2012) Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. The lancet 380:2163–2196
5. Charles A (2018) The pathophysiology of migraine: implications for clinical management. Lancet Neurol 17:174–182
6. Almalki ZA, Alzhrani MAG, Altowairqi AT et al (2018) Prevalence of migraine headache in Taif City, Saudi Arabia. J Clin Med Res 10:125
7. Latif R, Rafique N, Al Asoom L et al (2021) Diagnostic Accuracy of Serum Calcitonin Gene-Related Peptide and Apolipoprotein E in Migraine: A Preliminary Study. International Journal of General Medicine 14:851
8. Al Asoom LI, Alajmi MS, Alsudairi RR et al (2021) Association between sex hormones and migraine in young Saudi females. Saudi Med J 42:793–798
9. Al Jumah M, Al Khathaami AM, Kojan S et al (2020) The prevalence of primary headache disorders in Saudi Arabia: a cross-sectional population-based study. J Headache Pain 21:1–9
10. Dodick DW (2018) A phase-by-phase review of migraine pathophysiology. Headache: the journal of head face pain 58:4–16
11. Lipton RB, Bigal ME (2005) The epidemiology of migraine. The American Journal of Medicine Supplements 118:3–10
12. Samuel Appadurai Alexander R (2014) Study of Migraine Headache Spectrum in patients treated in the Tertiary Care Hospital. Stanley Medical College, Chennai
13. Anttila V, Stefansson H, Kallela M et al (2010) Genome-wide association study of migraine implicates a common susceptibility variant on 8q22. 1. Nat Genet 42:869
14. Gormley P, Anttila V, Winsvold BS et al (2016) Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. Nat Genet 48:856–866
15. Chang X, Pellegrino R, Garifallou J et al (2018) Common variants at 5q33. 1 predispose to migraine in African-American children. J Med Genet 55:831–836
16. Jiang Z, Zhao L, Zhang X et al (2021) Common variants in KCNK5 and FHL5 genes contributed to the susceptibility of migraine without aura in Han Chinese population. Sci Rep 11:1–8
17. van Hemert S, Breedveld AC, Rovers JM et al (2014) Migraine associated with gastrointestinal disorders: review of the literature and clinical implications. Front Neurol 5:241

18. Luo X, Kranzler HR, Zuo L et al (2006) ADH4 gene variation is associated with alcohol dependence and drug dependence in European Americans: results from HWD tests and case–control association studies. Neuropsychopharmacology 31:1085–1095
19. Sapio MR, Vessaz M, Thomas P et al (2015) Novel carboxypeptidase A6 (CPA6) mutations identified in patients with juvenile myoclonic and generalized epilepsy. PLoS One 10:e0123180
20. Weidinger S, Illig T, Baurecht H et al (2006) Loss-of-function variations within the filaggrin gene predispose for atopic dermatitis with allergic sensitizations. Journal of Allergy Clinical Immunology 118:214–219
21. Lipton RB, Bigal ME, Ashina S et al (2008) Cutaneous allodynia in the migraine population. Ann Neurol 63:148–158
22. Lee Y-C, Zhou Q, Chen J, Yuan J (2016) RPA-binding protein ETAA1 is an ATR activator involved in DNA replication stress response. Curr Biol 26:3257–3268
23. Wirth KG, Wutz G, Kudo NR et al (2006) Separase: a universal trigger for sister chromatid disjunction but not chromosome cycle progression. J Cell Biol 172:847–860
24. Sheng S, Carey J, Seftor EA et al (1996) Maspin acts at the cell membrane to inhibit invasion and motility of mammary and prostatic cancer cells. Proceedings of the National Academy of Sciences 93:11669–11674
25. Brandt DT, Baarlink C, Kitzing TM et al (2009) SCAI acts as a suppressor of cancer cell invasion through the transcriptional control of β 1-integrin. Nat Cell Biol 11:557–568
26. Schwedt TJ, Berisha V, Chong CD (2015) Temporal lobe cortical thickness correlations differentiate the migraine brain from the healthy brain. PLoS One 10:e0116687

Figures

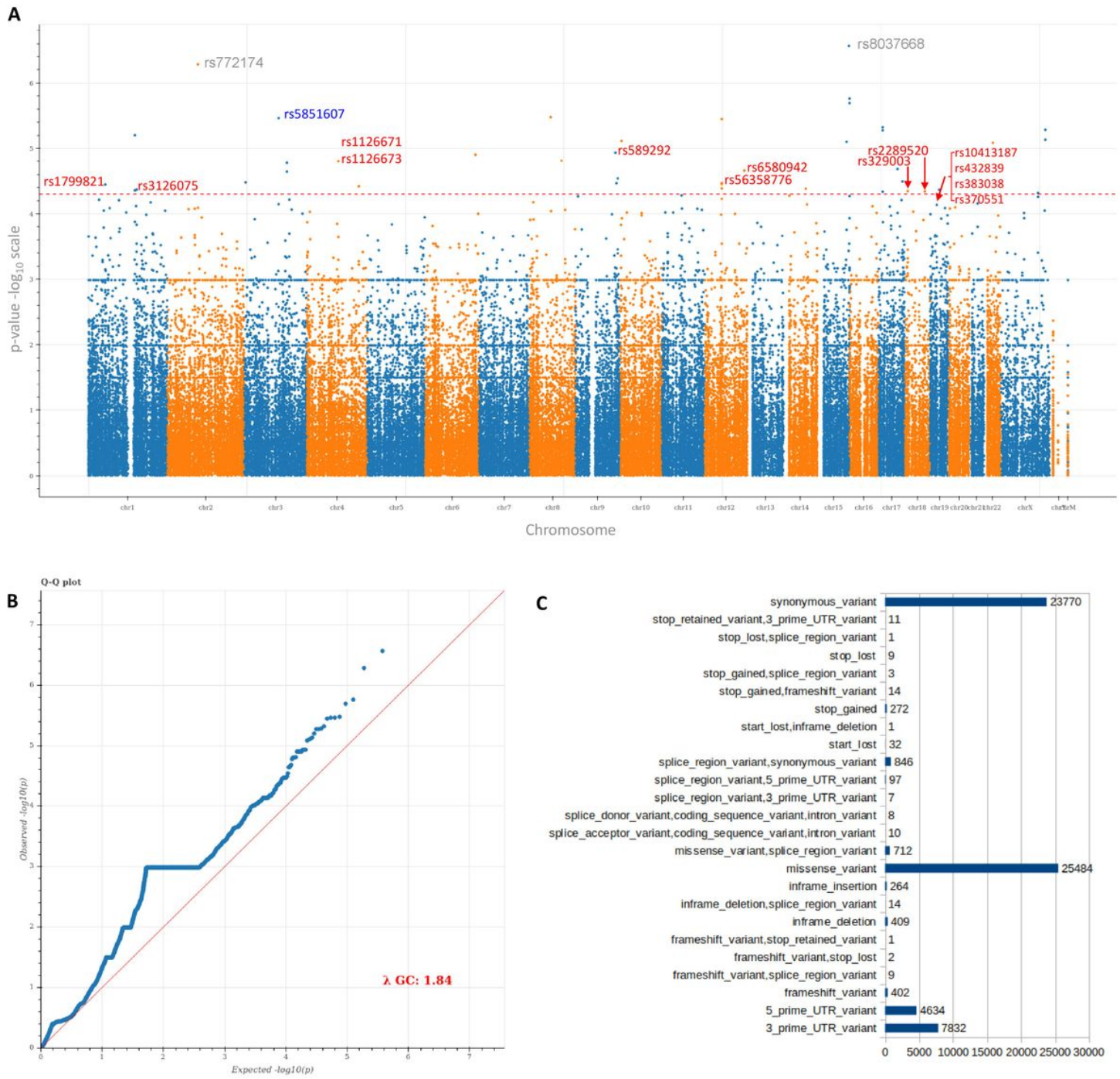

Figure 1

Manhattan plot (A) & QQ plot (B) of the association of SNPs with migraine as statistical significance in terms of p-values on a genomic scale. SNP numbers with blue color indicates the highly associated (p<3.44623×10-06) stop gained frameshift variant. SNP numbers with red color indicates the highly associated (p<0.00001) missense variants. SNP numbers with ash color indicates the highly associated (p<0.00001) intronic and 5 prime variants. Chr15: rs8037668 ['T', 'C'] (p value 2.7135×10-07) LOC400464 intron variant, non-coding transcript variant; and Chr 2: rs772174 ['A', 'G'] , (p-value 5.2006×10-07), ITPRIPL1 protein coding 5 prime UTR variant showed the highest p-value. C: Consequences of coding variants. Full list of variants (p value <0.00001) are listed in the Table S1.

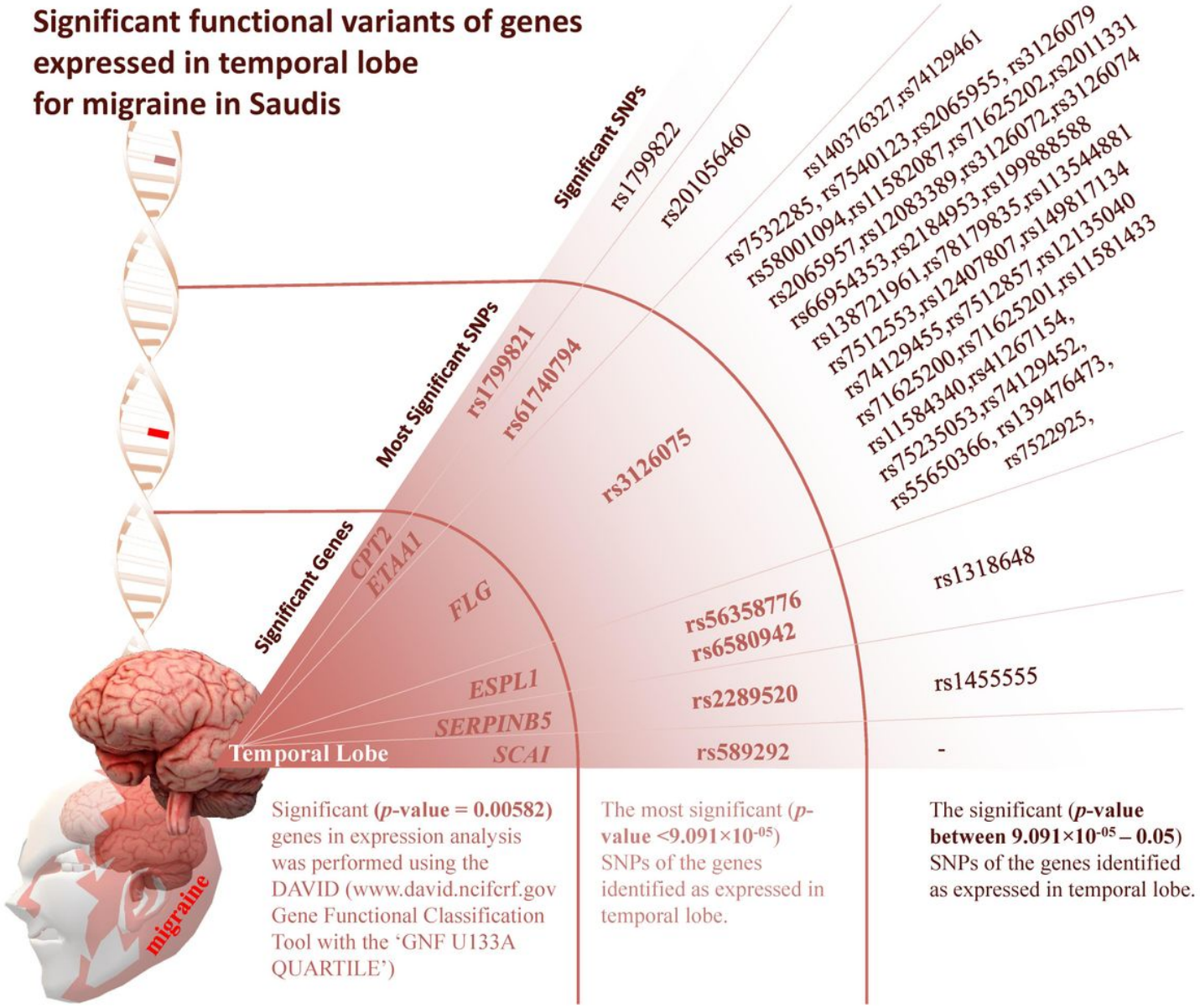

Figure 2

Significant functional variants of genes expressed in temporal lobe (p-value = 0.00582) for migraine in Saudis. The most significant genes (top 12 as listed in the table 3) associated functionally were used as input to identify the expression nature of them using the gene functional classification tool, DAVID with the 'GNF U133A QUARTILE'. The most significant (p-value <9.091×10-05) functional variants are presented in the middle path. The functional variants with p-value between 9.091×10-05 to 0.05 on the genes identified as expressed in temporal lobe are also presented. A total of 37 functional variants are found to be significant in the gene, FLG (Full list presented in Table 4).

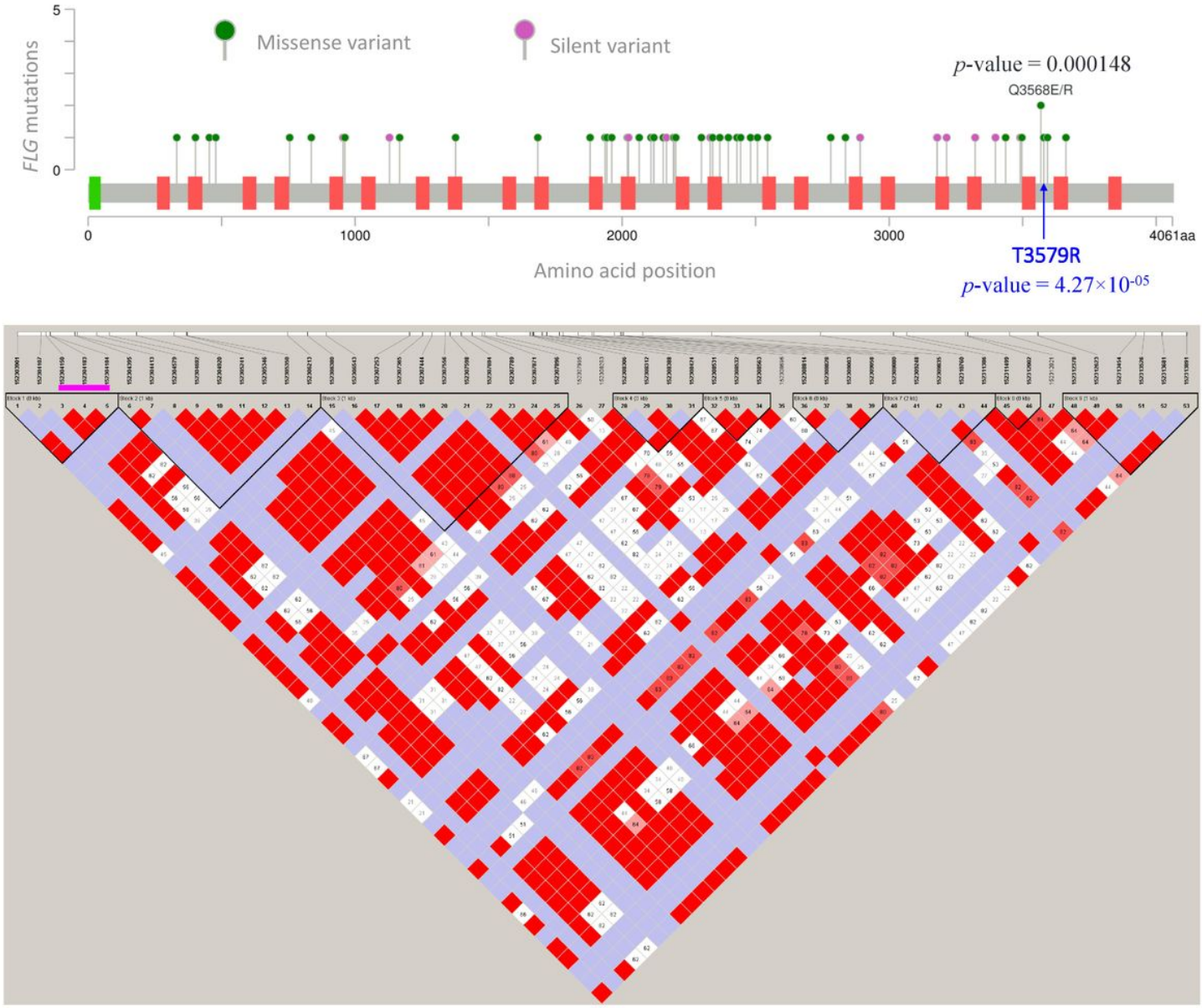

Figure 3

Top: Variants observed in FLG gene. Pink lollipop indicates silent mutation; Green lollipop indicates missense variations. Amino acid substitution in blue color indicates the most significant variant. Box colored green indicates calcium binding domain. Bottom: Haplomap of gene level linkage disequilibrium analysis of the SNPs in the FLG gene. Pink line: The most significant three markers rs3126075G, rs7532285T and rs7540123G.

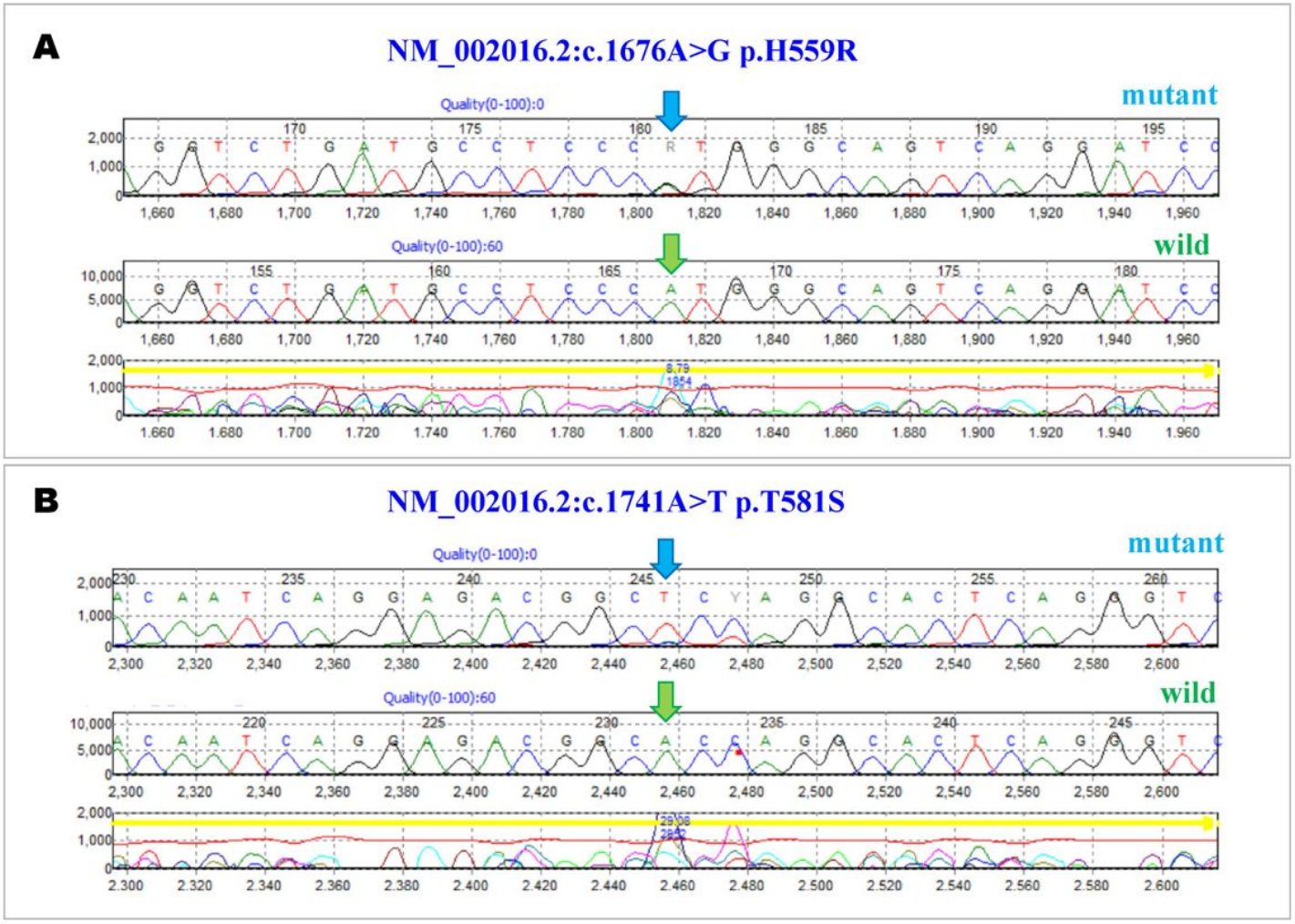

Figure 4

Electropherogram of FLG gene with (A): NM_002016.2:c.1676A>G p.H559R and (B): NM_002016.2:c.1741A>T p.T581S. Blue arrow indicates mutant. Green arrow indicates wild.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigureS1hsa04115p53Signallingpathway.png
- FigureS2hsa00071fattyAcidDegradation.png
- FigureS3hsa05322Systemiclupuserythematosus.png
- Supplementarytables.docx