BIOI4464 Course Project: The LPL gene

The LPL gene & protein:

LPL gene is located at the plus strand of the 8th chromosome (8p21.3) stretching from locus chr8:19,901,717 to 19,967,259 making the gene size of 65,543 bases. The lipoprotein lipase produced by the LPL gene has a length of 475 amino acids (table 1) and is a crucial factor in lipid metabolism. The LPL protein is found in nearly all tissues, but is especially enhanced in adipose tissue, breast tissue and heart muscle. (Genecards: GC08P019901, UniProtKB: P06858)

```
>sp|P06858|LIPL_HUMAN Lipoprotein lipase OS=Homo sapiens OX=9606 GN=LPL PE=1 SV=1
MESKALLVLTLAVWLQSLTASRGGVAAADQRRDFIDIESKFALRTPEDTAEDTCHLIPGV
AESVATCHFNHSSKTFMVIHGWTVTGMYESWVPKLVAALYKREPDSNVIVVDWLSRAQEH
YPVSAGYTKLVGQDVARFINWMEEEFNYPLDNVHLLGYSLGAHAAGIAGSLTNKKVNRIT
GLDPAGPNFEYAEAPSRLSPDDADFVDVLHTFTRGSPGRSIGIQKPVGHVDIYPNGGTFQ
PGCNIGEAIRVIAERGLGDVDQLVKCSHERSIHLFIDSLLNEENPSKAYRCSSKEAFEKG
LCLSCRKNRCNNLGYEINKVRAKRSSKMYLKTRSQMPYKVFHYQVKIHFSGTESETHTNQ
AFEISLYGTVAESENIPFTLPEVSTNKTYSFLIYTEVDIGELLMLKLKWKSDSYFSWSDW
WSSPGFAIQKIRVKAGETQKKVIFCSREKVSHLQKGKAPAVFVKCHDKSLNKKSG
```

Table 1. LPL protein amino acid sequence (UniProtKB: P06858)

The canonical transcript of the LPL protein in Ensembl (LPL-207 ENST00000650287.1) is a 1:1 match to the sequence above. The Ensembl gives following information considering the transcript: is annotated with 33 domains and features, is associated with 8782 variant alleles and maps to 561 oligo probes.

Once synthesized, the LPL protein attaches to heparan sulfate proteoglycans (HSPGs). With the help of HSPGs the LPL reaches the glycosylphosphatidylinositol (GPI)-anchored high-density lipoprotein–binding protein 1 (GPIHBP1) and a complex of LPL–GPIHBP1 is formed. The complex mediates from parenchymal cells to the surface of capillary endothelium and where it attaches and allows the complex to catch chylomicrons. Once the chylomicron is caught, the triglycerides are extracted and hydrolyzed by the LPL producing diglyceride and one fatty acid. (Birrane et al. 2019; Horton 2019; UniProtKB: P06858)

The mutations in LPL are known to cause problems with the triglyceride metabolism and can lead to malfunction of the LPL and to the familial chylomicronemia syndrome where triglyceride levels rise when feeding on a normal diet. The rise of the triglyceride levels is a result of the accumulated chylomicron and the situation can lead to pancreatitis, fatigue, various GI symptoms, hepatosplenomegaly, lipid accumulation to skin bulges (eruptive xanthomas), neurological disorders and increase a risk to familial combined hyperlipidemia. (OMIM: 609708, ORPHA:444490)

Conserved sites of the LPL protein:

The InterPro classificatory classifies the LPL (P0658) to belong into two superfamilies: Alpha/Beta hydrolase fold (IPR029058) and PLAT/LH2 domain superfamily (IPR036392). Furtherly the Alpha/Beta hydrolase fold superfamilie includes the subfamilies in descending order: triacylglyserol (IPR000734) → Lipase, LIPH-type (IPR016272) → Lipoprotein lipase (IPR002330). Following the Alpha/Beta branch the InterPro classifies LPL having two sub domains: Lipase/vitellogenin (IPR013818) → Lipase, N-terminal (IPR033906).

The InterPro classifies the LPL (P0658) to have the following subdomain under the PLAT/LH2 domain superfamily: PLAT/LH2 domain (IPR001024).

Alpha/Beta hydrolase fold (IPR029058) located between amino acids 23-339 is a 3D structure of the protein in which eight beta strands are connected by alpha helices. The structure is common with proteins associating to hydrolyze and the most conserved part of the structure is the catalytic triad located at the

loop. The catalytic triad of the LPL - titled as the nucleophile elbow - associates to the hydrolysation of the triglycerides and is the most conserved feature of the LPLs alpha/beta hydrolase fold. The triad is formed under serine residue together with histidine and aspargine residues. The preservations of the triad is easily seen in the Conserved Domains and Protein Classification database (CDD/SPARCLE) where the preservations is seen perfectly between species (cd00707).

As mentioned the alpha/beta hydrolase and moreover the catalytic triad are common in lipases as can be seen from the CDD/SPARCLE database (cd00741). UniProtKB lists 129 reviewed human lipases to exist (lipase AND organism:"Homo sapiens (Human) [9606]") and InterPro classifies closest relatives to be pancreatic and hepatic lipases. Few picks from the lipase entry of CDD/SPARCLE (cd00741): Diacylglycerol lipase beta, phospholipase A1 member A isoform 1 precursor, membrane-associated phospholipase A1 beta.

PLAT/LH2 domain (IPR001024) located between amino acids 341-465 is associated with various lipid or membrane proteins. PLAT/HL2's 3D structure is a sandwich like two folded beta strand having two opposite sheets with four beta strands in each sheet. The domains structure makes it a highly conserved region with residues located at the core parts of the protein. The exception being lysine or argine located at the surface of the fifth beta-strand of the eukaryotic domains. The PLAT/HL2 domain is crucial for the protein-protein interactions to membrane bound proteins. The GPIHBP1 that intermediates with LPL is bound to the region of the PLAT/HL2 domain (Birrane et al. 2019).

The PLAT_lipase seen from the CDD/SPARCLE (cd01755) has the conserved domain similar to Hepatic triacylglycerol lipase, Endothelial lipase, Chain A Triacylglycerol Lipase Pancreatic and computationally predicted similar to lipase CoPL-RP2. Other relatives can be expected to find if zoomed out furtherly to the PLAT/LH2 domain superfamily level.

SNPs affecting the LPL protein:

There are plenty of variations affecting the LPL gene and several of them can lead to the deficiency and/or malfunction of the LPL protein. The UniprotKB lists of total 146 variation and 17 mutations affecting to LPL protein. According to UEF genome center the mutation affecting to the LPL protein lie particularly at the exon 5 in codons 176, 188, 194, 205 ja 207. The Ensembl database lists SNPs for LPL a staggering 41674 pieces and somatic SNVs 1231 mutations, fortunately non-synonymous missense SNPs are "only" 796 pieces (Ensembl: LPL ENSG00000175445). Also the company Blueprint Genetics has the LPL gene in its selections and lists two non-coding disease causing variants rs540525285 and rs328 that are covered with their test (https://blueprintgenetics.com/tests/single-gene-tests/lpl-single-gene-test-2/).

To reduce the number of hits, the results for missense SNPs were downloaded and formatted as an excel file. A filtering according to the clinical significance was done and only those SNPs that are pathogenic, likely pathogenic or uncertain significance pathogenic risk factor were accepted to the list. The procedure produced a list of 40 SNPs that was furtherly modified so that duplications were removed from the list, the result was a list of 28 SNPs listed in the table 1 and organized by the coordination of the changed AA.

The results are categorized using Polyphen value (1 most damaging, 0 bening). The SIFT value was also left for comparison and seems to follow quite well the Polyphen value, although few differences occur. Some SNPs are categorized by SIFT and/or Polyphen categorize as benign or tolerated but were left to the table as the clinical significance is marked as pathogenic. (Ensembl: LPL ENSG00000175445; UEF Genome center)

Unaware of when and where has the UEF Genome center spotted the mutations, I looked at exon 5 (figure 1) in the Ensembl and found two SNPs that are pathogenic and located near the area mentioned: the rs118204056 Ala203Thr and rs118204076 Asp207Glu, both are in the table 1. Either rs540525285 or rs328

weren't classified to pathogenic by the Ensembl but for curiosity I decided to add those to the table 2 to be taken for the VEP analyze.

#	Variant ID	Location	vf_allele	Alleles	Clin. Sig.	Conseq. Type	AA	AA coord	sift_class	SIFT	polyphen_class	PolyPhen
1	rs118204073	8:19951825	C	A/C	pathogenic	missense variant	R/S	26	tolerated	0.15	probably damaging	0.995
2	rs118204069	8:19951856	C	T/C	pathogenic	missense variant	W/R	37	deleterious	0	probably damaging	1
3	rs118204058	8:19951916	G	C/G/T	pathogenic	missense variant	Q/E	57	tolerated	0.33	benign	0.003
4	rs118204063	8:19953386	A	G/A	pathogenic	missense variant	G/E	93	deleterious	0	probably damaging	1
5	rs118204064	8:19954126	G	A/G	pathogenic	missense variant	D/G	107	deleterious	0	probably damaging	1
6	rs372668179	8:19954168	T	G/A/T	pathogenic	missense variant	R/L	197	deleterious	0	possibly damaging	0.817
7	rs118204072	8:19954174	G	C/G/T	pathogenic	missense variant	S/C	199	deleterious	0	probably damaging	1
8	rs118204056	8:19954185	A	G/A	pathogenic	missense variant	A/T	203	deleterious	0	probably damaging	1
9	rs118204076	8:19954199	G	C/G/T	pathogenic	missense variant	D/E	207	deleterious	0	probably damaging	1
10	rs118204057	8:19954222	A	G/A/C	pathogenic	missense variant	G/E	215	tolerated	0.43	probably damaging	0.999
11	rs118204061	8:19954240	C	T/C	pathogenic	missense variant	I/T	221	deleterious	0	probably damaging	1
12	rs118204075	8:19954243	A	G/A	pathogenic	missense variant	G/E	222	deleterious	0	probably damaging	1
13	rs118204067	8:19954271	G	C/G	pathogenic	missense variant	D/E	231	deleterious	0	probably damaging	0.911
14	rs118204060	8:19954279	T	C/T	pathogenic	missense variant	P/L	234	deleterious	0	probably damaging	1
15	rs118204080	8:19954333	C	T/C	pathogenic	missense variant	I/T	252	deleterious	0.02	possibly damaging	0.461
16	rs1554517725	8:19955862	A	G/A	likely pathogenic	missense variant	C/Y	266	deleterious	0	probably damaging	1
17	rs118204082	8:19955863	G	C/G/T	pathogenic	missense variant	C/W	266	deleterious	0	probably damaging	1
18	rs118204077	8:19955873	T	C/T	pathogenic	missense variant	R/C	270	deleterious	0	probably damaging	1
19	rs118204062	8:19955874	A	G/A	pathogenic	missense variant	R/H	270	deleterious	0	probably damaging	1
20	rs118204059	8:19955876	A	T/A	pathogenic	missense variant	S/T	271	deleterious	0	probably damaging	0.947
21	rs118204068	8:19955894	A	G/A	pathogenic	missense variant	D/N	277	deleterious	0.03	benign	0.44
22	rs1064797075	8:19955969	C	T/C	likely pathogenic	missense variant	C/R	302	deleterious	0	possibly damaging	0.9
23	rs886037774	8:19955993	C	T/C	likely pathogenic	missense variant	C/R	310	deleterious	0	probably damaging	0.993
24	rs268	8:19956018	G	A/G	uncertain significanc	missense variant	N/S	318	tolerated	0.24	benign	0.137
25	rs118204071	8:19959322	A	G/A	pathogenic	missense variant	A/T	361	tolerated	1	benign	0.142
26	rs118204078	8:19960935	G	C/G	pathogenic	missense variant	L/V	392	deleterious	0	probably damaging	0.999
27	rs886037775	8:19960948	T	A/T	likely pathogenic	missense variant	E/V	396	deleterious	0	probably damaging	0.994
28	rs118204079	8:19962126	A	G/A	pathogenic	missense variant	C/Y	445	deleterious	0	probably damaging	0.961
29	rs328	8 :19962213	C	C/G	benign/likely ben.	stop gained	S/Y	474	.	.	.	.
30	rs540525285	8:19939200	C	G/C	.	intron variant	.	.	.	.	.	.

Table 2. SNPs for VEP, filtered by the clinical significance.

No.	Exon / Intron	Start	End	Start Phase	End Phase	Length
5	ENSE00001206556	19,954,120	19,954,353	1	1	234

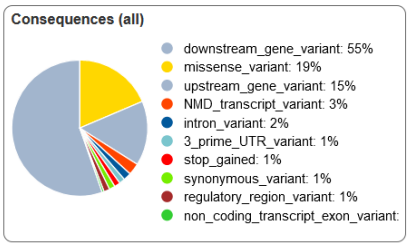
Sequence

GCCTCGATCCAGCTGGACCTAACTTTCAGTATGCAGAAGCCCCGAGTCGTCTTTCTCCTG
ATGATGCAGATTTTGTAGACGTCTTACACACATTCACCAGAGGGTCCCCTGGTCGAAGCA
TTGGAATCCAGAACCAGTTGGGCATGTTGACATTTACCCGAATGGAGGTACTTTTCAGC
CAGGATGTAACATTGGAGAAGCTATCGCGTGATTGCAGAGAGGACTTGGAG

Figure 1. Ensembl entry for the LPL's exon 5.

As mentioned, the LPL gene has tehns of thousans mutations affecting to it, UniProtKB lists hundreds of them and OMIM 41 allelic variants for the LPL protein (OMIM: 609708). For simplicity reasons the VEP analyze was done with only those variants shown in the table 1. Variant id's (starting with rs) were copy pasted to the Ensembl Variant Effect Predictor web service. Aditional adjustments were done so that protein name, exon and intron numbers and protein domane columns were selected to be shown. A summary of the VEP analyze is shown in the figure 2.

Category	Count
Variants processed	30
Variants filtered out	0
Novel / existing variants	0 (0.0) / 30 (100.0)
Overlapped genes	1
Overlapped transcripts	8
Overlapped regulatory features	2



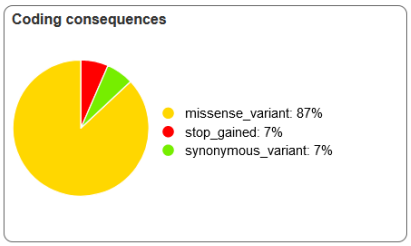


Figure 2. VEP analyze of the SNPs

Only gene that was affected by the variants was the LPL gene. The most serious consequences like complete deletions transcript_ablation or splice_acceptor_variant are not seen in the results. The consequences caused by the SNP are mainly missense variants and considered to have only moderate impact. The stop_gained is considered as high impact and synonymous - where there is no resulting change to the encoded amino acid - as low. The uploaded variants have an effect to many of the observed transcripts of the LPL gene, but to simplify the results were filtered so that only the canonical transcript ENST00000650287.1 is visible (table 3).

Based on the VEP results, only exons 1-2 and 10 aren't affected by the selected variants. The most prone exons for the selected SNPs are the exons 5 and 6. The result for exon 5 is in accordance with the UEF Genome centers information that the exon 5 is one of the most mutation prone areas of the LPL gene.

#Uploaded_variation	Location	Allele	Consequence	IMPACT	Exon	AA	SIFT	PolyPhen
rs540525285	8:19939200-19939200	C	upstream_gene_variant	MODIFIER	.	-	-	-
rs118204073	8:19951825-19951825	C	missense_variant	MODERATE	3/10	R/S	deleterious(0)	probably_damaging(0.995)
rs118204069	8:19951856-19951856	C	missense_variant	MODERATE	3/10	W/R	deleterious(0)	probably_damaging(1)
rs118204058	8:19951916-19951916	G	missense_variant	MODERATE	3/10	Q/E	tolerated(0.3)	benign(0.003)
rs118204058	8:19951916-19951916	T	stop_gained	HIGH	3/10	Q/*	-	-
rs118204063	8:19953386-19953386	A	missense_variant	MODERATE	5/10	G/E	deleterious(0)	probably_damaging(1)
rs118204064	8:19954126-19954126	G	missense_variant	MODERATE	5/10	D/G	deleterious(0)	probably_damaging(1)
rs372668179	8:19954168-19954168	A	missense_variant	MODERATE	5/10	R/H	deleterious(0)	probably_damaging(0.988)
rs372668179	8:19954168-19954168	T	missense_variant	MODERATE	5/10	R/L	deleterious(0)	possibly_damaging(0.817)
rs118204072	8:19954174-19954174	G	missense_variant	MODERATE	5/10	S/C	deleterious(0)	probably_damaging(1)
rs118204072	8:19954174-19954174	T	missense_variant	MODERATE	5/10	S/F	deleterious(0)	probably_damaging(1)
rs118204056	8:19954185-19954185	A	missense_variant	MODERATE	5/10	A/T	deleterious(0)	probably_damaging(1)
rs118204076	8:19954199-19954199	G	missense_variant	MODERATE	5/10	D/E	deleterious(0)	probably_damaging(1)
rs118204076	8:19954199-19954199	T	synonymous_variant	LOW	5/10	D	-	-
rs118204057	8:19954222-19954222	A	missense_variant	MODERATE	5/10	G/E	tolerated(0.43)	probably_damaging(0.999)
rs118204057	8:19954222-19954222	C	missense_variant	MODERATE	5/10	G/A	deleterious(0.03)	probably_damaging(0.999)
rs118204061	8:19954240-19954240	C	missense_variant	MODERATE	5/10	I/T	deleterious(0)	probably_damaging(1)
rs118204075	8:19954243-19954243	A	missense_variant	MODERATE	5/10	G/E	deleterious(0)	probably_damaging(1)
rs118204067	8:19954271-19954271	G	missense_variant	MODERATE	5/10	D/E	deleterious(0)	probably_damaging(0.911)
rs118204060	8:19954279-19954279	T	missense_variant	MODERATE	5/10	P/L	deleterious(0)	probably_damaging(1)
rs118204080	8:19954333-19954333	C	missense_variant	MODERATE	5/10	I/T	deleterious(0.02)	possibly_damaging(0.461)
rs1554517725	8:19955862-19955862	A	missense_variant	MODERATE	6/10	C/Y	deleterious(0)	probably_damaging(1)
rs118204082	8:19955863-19955863	G	missense_variant	MODERATE	6/10	C/W	deleterious(0)	probably_damaging(1)
rs118204082	8:19955863-19955863	T	synonymous_variant	LOW	6/10	C	-	-
rs118204077	8:19955873-19955873	T	missense_variant	MODERATE	6/10	R/C	deleterious(0)	probably_damaging(1)
rs118204062	8:19955874-19955874	A	missense_variant	MODERATE	6/10	R/H	deleterious(0)	probably_damaging(1)
rs118204059	8:19955876-19955876	A	missense_variant	MODERATE	6/10	S/T	deleterious(0)	probably_damaging(0.947)
rs118204068	8:19955894-19955894	A	missense_variant	MODERATE	6/10	D/N	deleterious(0.03)	benign(0.44)
rs1064797075	8:19955969-19955969	C	missense_variant	MODERATE	6/10	C/R	deleterious(0)	possibly_damaging(0.9)
rs886037774	8:19955993-19955993	C	missense_variant	MODERATE	6/10	C/R	deleterious(0)	probably_damaging(0.993)
rs268	8:19956018-19956018	G	missense_variant	MODERATE	6/10	N/S	tolerated(0.24)	benign(0.137)
rs118204071	8:19959322-19959322	A	missense_variant	MODERATE	7/10	A/T	tolerated(1)	benign(0.142)
rs118204078	8:19960935-19960935	G	missense_variant	MODERATE	8/10	L/V	deleterious(0)	probably_damaging(0.999)
rs886037775	8:19960948-19960948	T	missense_variant	MODERATE	8/10	E/V	deleterious(0)	probably_damaging(0.994)
rs118204079	8:19962126-19962126	A	missense_variant	MODERATE	8/10	C/Y	deleterious(0)	probably_damaging(0.961)
rs328	8:19962213-19962213	G	stop_gained	HIGH	9/10	S/*	-	-

Table 3. VEP results for transcript ENST00000650287.1

OMIM lists 43 variants affecting the LPL gene and out of them 36 has a rs ID that can be compared to the VEP analyze results. Out of the 36 OMIM entries, 24 are included at the VEP analyze nd some Ensembl VEP variants aren't found at the OMIM (table 4).

OMIM (rs#)	VEP(rs#)		OMIM (rs#)	VEP(rs#)
268	268		118204072	118204072
326	NA		118204073	118204073
328	328		118204074	NA
13702	NA		118204075	118204075
1801177	NA		118204076	118204076
118204056	118204056		118204077	118204077
118204057	118204057		118204078	118204078
118204058	118204058		118204079	118204079
118204059	118204059		118204080	118204080
118204060	118204060		118204081	NA
118204061	118204061		118204082	118204082
118204062	118204062		766134215	NA
118204063	118204063		1563569634	NA
118204064	118204064		1563572716	NA
118204065	NA		1563575252	NA
118204066	NA		NA	372668179
118204067	118204067		NA	540525285
118204068	118204068		NA	886037774
118204069	118204069		NA	886037775
118204070	NA		NA	1064797075
118204071	118204071		NA	1554517725

Table 4. OMIM vs. VEP analyze

Paralogs: BLAST & MSA

The retrieval of similar sequences was done with NCBI Blast using protein blast and limiting the search by organism (human: 9606) and database to reference proteins (ref_seq protein). Contrast to the whole protein sequence, the blastp search was done using only positions 28 – 475 which is annotated as the actual Lipoprotein lipase part of the protein, the BLASTed sequence is showm below (table 5). The decision to use the actual protein part was made to get more coherent results from the BLAST search, the blastp search resulted in total 35 hits and results are attached to appendix 1. In the results there is a major drop after hepatic triacylglycerol lipase isoform X1 at position 6 where the e-value drops from 1e-136 to hepatic triacylglycerol lipase isoform X1's 4e-99. After that the e-value declines quite steadily and the final hit phospholipase A1 member A isoform 3 has an e of 8e-24. All 35 hits were renamed accordingly to the NCBI nucleotide -database entries and taken for MSA.

```
>sp|P06858|28-475
ADQRRDFIDIESKFALRTPEDTAEDTCHLIPGVAESVATCHFNHSSKTFMVIHGWTVTGM
YESWVPKLVAALYKREPDSNVIVVDWLSRAQEHYPVSAGYTKLVGQDVARFINWMEEEFN
YPLDNVHLLGYSLGAHAAGIAGSLTNKKVNRITGLDPAGPNFEYAEAPSRLSPDDADFVD
VLHTFTRGSPGRSIGIQKPVGHVDIYPNGGTFQPGCNIGEAIRVIAERGLGDVDQLVKCS
HERSIHLFIDSLLNEENPSKAYRCSSKEAFEKGLCLSCRKNRCNNLGYEINKVRAKRSSK
MYLKTRSQMPYKVFHYQVKIHFSGTESETHTNQAFEISLYGTVAESENIPFTLPEVSTNK
TYSFLIYTEVDIGELLMLKLKWKSDSYFSWSDWWSSPGFAIQKIRVKAGETQKKVIFCSR
EKVSHLQKGKAPAVFVKCHDKSLNKKSG
```

Table 5. LPL for the BLAST.

Tha MSA for the found sequences was done in European Molecular Biology Laboratory - European Bioinformatics Institute (EMBL-EBI) web server and Clustal Omega was chosen to be the algorithm used. Following adjustments were done to the settings: number of iterations 5, max guide tree iterations 5, max HMM iterations 5. Other settings were left as default. The visualization of the first MSA was done with GeneDoc in where the MSA from the EMBL-EBI was opened and no other changes than taking of the similarity groups was done. The clustal formatted MSA is as an appendix 2.

The BLAST search included only 10 RefSeq Selected genes (LPL, LIPG, LIPC, LIPH, LIPI, PLA1A, PNLIP, PNLIPRP1, PNLIPRP2, PNLIPRP3) and the final result (35) includes several variants of the 10 genes. As can be seen from the appendix 1, the isoforms are of several lengths and this was expected to have an effect to the MSA results. On inspected, the first MSA looks quite heterogeneous having some long middle gaps and the otherwise well preserved blocks are cancelled by a gap of some significantly shorter protein. Also the percent identity matrix (PIM) in appendix 3 indicates several transcript variants being nearly identical or identical. This indicated the need for pruning the dataset to get more coherent results.

The pruning was done with Seaview by manually inspecting sequences and deleting highly similar sequences (as seen in PIM), truncated or otherwise misaligned sequences. In case of highly similar sequences, the sequence that was used as a RefSeg in NCBI was selected. The clustal omega alignment from EMBL-EBI was downloaded and transformed to .fasta format with ClustalX 2.1 after which the sequence was saved as an FASTA format. The .fasta file formed was then opened in Seaview where all above mentioned pruning was done.

The pruned alignment (appendix 4) contained only the main variants of the aforementioned genes, other isoforms were considered either misaligns, near duplicates or uninformative due to alternative splicing. The cleaned MSA was used as a basis of the final MSA that was done in the EMBL-EBI Clustal O and following adjustments were done to the parameters: max guide tree iterations 5, max HMM iterations 5, number of combined iterations 5, order input. The clustal formatted MSA is as an appendix 5.

The final MSA shows strong preserved clusters especially in the middle parts (starting from G at position 132) of the sequences. Could this be the Alpha/Beta hydrolase fold part? All in all the alignment looks decent, though there are some ugly caps in some middle sections and sequences of PLA1A, LIPH and LIPI are shorter than the others.

Comparison of protein MSA and nucleotide MSA in tree building

The final MSA was used in building two phylogenetic trees based on the amino acid sequences and DNA sequences. The cDNA sequences of the selected genes were manually retrieved from the Ensembl in where the transcript table was used to select the main transcript of the gene. The selected transcripts are listed below in table 6.

>ENST00000299022.10 LIPC-201 cdna:protein_coding
>ENST00000650287.1 LPL-207 cdna:protein_coding
>ENST00000261292.9 LIPG-201 cdna:protein_coding
>ENST00000369230.4 PNLIPRP3-201 cdna:protein_coding
>ENST00000591655.3 PNLIPRP2-204 cdna:protein_coding
>ENST00000369221.2 PNLIP-201 cdna:protein_coding
>ENST00000358834.9 PNLIPRP1-201 cdna:protein_coding
>ENST00000273371.9 PLA1A-201 cdna:protein_coding
>ENST00000296252.9 LIPH-201 cdna:protein_coding
>ENST00000344577.6 LIPI-201 cdna:protein_coding

Table 6. Ensembl's cDNA entries for the selected genes.

The cDNA sequences were aligned in pal2nal server using the protein MSA as a template. The cDNA sequences were in the same order as the protein sequences and both sequences were uploaded to the server in .fasta -format. Option settings were following: Codon table/Universal code, Remove gaps inframe

stop codons/No, Remove mismatches: (mismatched codons between protein and DNA)/Yes, Use only selected positions ('#' under the input alignment)/No, Output format/Clustal.

The MSAs of cDNA and protein sequences were imported in the MEGA-X software and furtherly transformed into .mega file for the MEGA-X to handle them better. A manual inspection was done to both sequences to see the quality of the alignments and based on the inspection gaps/missing data treatment was decided to be partial and cutoff 30%. The decision was based on observation that there were many parts in the alignment where only 1-3 sequences were aligned against gaps and that those parts where 50% or more sequences were preserved were thought to be interesting.

For the tree construction the maximum likelihood method was chosen as it seems to be the best non Bayesian approach to construct a robust phylogenetic tree. For test of phylogeny a bootstrap method was chosen so that the tree quality could be more rigorously estimated. (Douady et al. 2003; Holder & Lewis 2003).

A standard method is to do several thousand repeats for bootstrapping and in this case a number of replications was decided to be 5000. This was estimated to produce enough repeats yet also be time efficient. At first no attention was paid to the substitution models and they were let to default chosen by MEGA-X; for protein sequences the model was Jones-Taylor-Thornton model and to cDNA the model was Tamura-Nei. The results for the 5000 bootstraps and 30% partial deletion are below in figures 3 and 4.

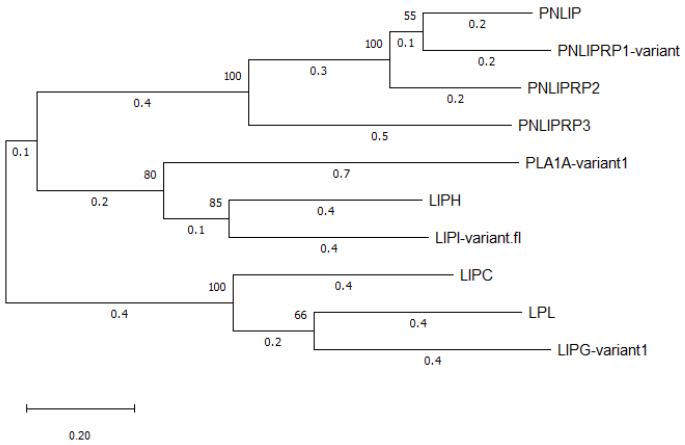


Figure 3. Protein based phylogenetic tree for the LPL like sequences (JTT).

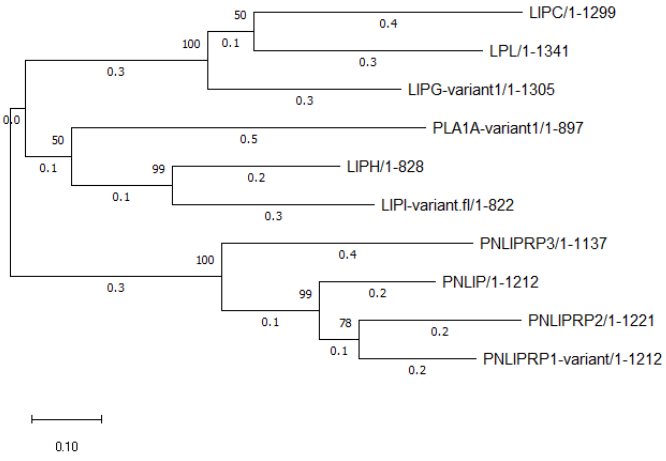


Figure 4. cDNA based phylogenetic tree for the LPL like sequences (Tamura-Nei).

After trying to get more sense to the substitution models and essentially not getting any sense, I run into guide for MEGA5 and saw a mention that the MEGA can estimate the best substitution model for the given sequences (Hall 2013). I decided to have a look at the MEGA-X and indeed found that the X version had exactly same feature for model estimation that I decided to try.

For both the protein and cDNA sequences automatic model selection was used: Models→Find best DNA/Protein models (ML) with default settings described in figure 5. The figure is from estimating best model for the amino acid sequences but estimation was done in similar manner for cDNA.

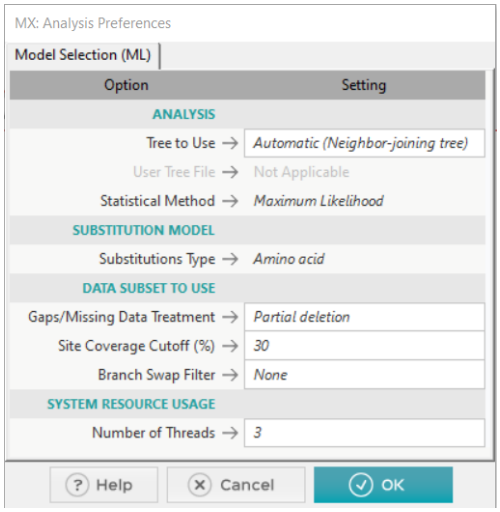


Figure 5. Settings for model testing.

The analyze produced a list in where the models with ranked with Bayesian Information Criterion scores best to worst (the lower the better). MEGA-X recommended for the cDNA a K2+G+I method (20 parameters, BIC score 24357.009 and for protein WAG+G+I (19 parameters, BIC score 14364.515). Using the same MSAs and same number of repeats and same deletion cutoff a rerun was done using WAG model with gamma distribution (+G) 5 rate categories and evolutionarily invariable (+I) and for the cDNA K2 (or K80 or K2P) model with G and I.

For WAG model see Whelan & Goldman 2001, for K2 see Kimura 1980. The reanalyzed trees are below in figures 6 and 7.

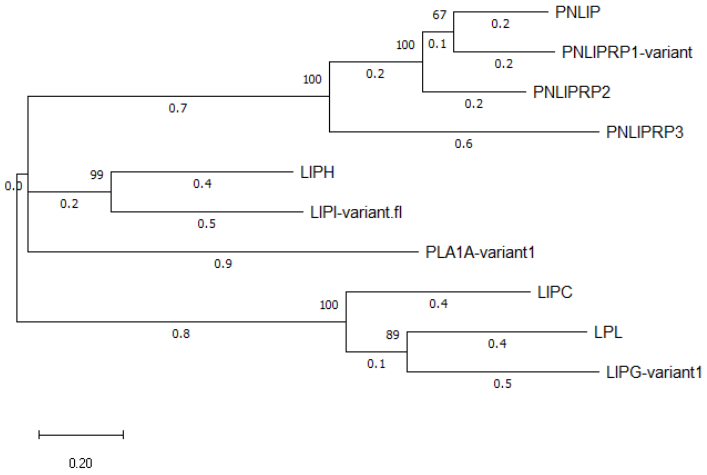


Figure 6. WAG+G+I tree for protein sequences

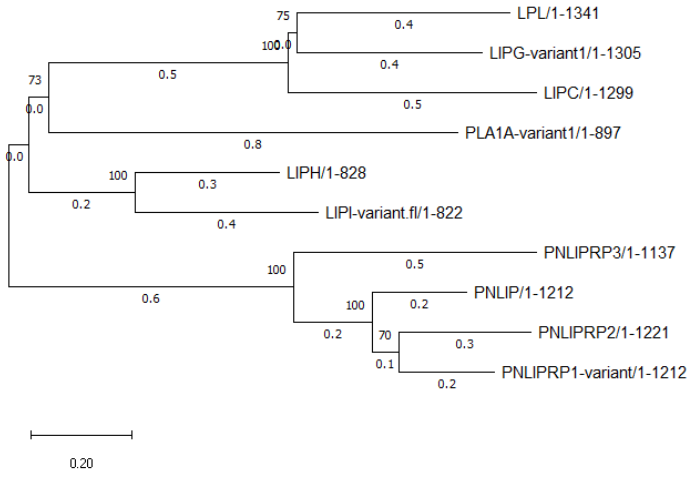


Figure 7. K2+G+I tree for cDNA sequences.

Both cDNA and protein trees and both substitution models are consistent in categorizing the genes. A slight difference is how the PNLIP is ordered (in protein trees together with PNLIRP1, in cDNA trees outside of the PNLIRP1 & 2 but in the same node) but this is logical the order of the genes can vary in the nodes within the given probability. Both protein and cDNA seem to categorize PLA1A to be the most distinctive of the LPL like genes.

The main difference in using the “correct” models is that the probabilities got better (trees more robust?) and there aren’t any more probabilities near 50%. A thing to notice in the “correct” models is that the branch lengths (evolutionary distances?) of the LIP and PNLIP groups have changed so that the lengths from the root have grown (more time from the split of the groups?). In contrast to other trees in the K2 cDNA tree the position of the PLA1A has changed to be diverged from the LIPH/LIPI group to LPL/LIPG/LIPC group.

Discussion:

The databases were selected on bases of familiarity to the author and also by a common reputation. For this work, I couldn’t find any reviews where the reliability of Alliance/Ensembl/ EMBL-EBI/NCBI/Uniprot/Omim/Orphanet databases would have been investigated. But as much the author is concerned, all of the mentioned databases should have an excellent reputation and they are curated/maintained regularly.

For BLAST searches the NCBI blast was used as it was part of the exercises in the course and is an intuitive site. The MSA was done in the EMBL-EBI site as the site offers free and easy to use MSA methods.

The ClustalΩ was chosen to be the MSA method as it seems to be if not the golden method, but at least quite common and fast/reliable algorithm (figure 8) for doing MSAs. Also the iterations were chosen to be done as many times as possible as it is possible that the iterations improve the accuracy of the MSA. (Le et al. 2017; Sievers & Higgins 2018)

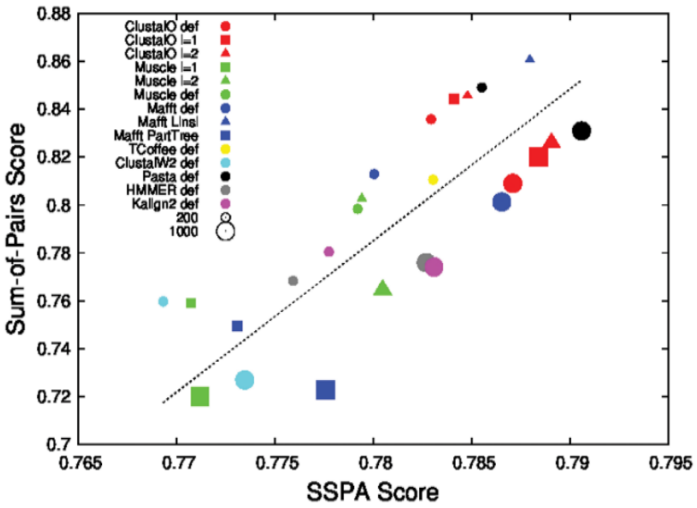


Figure 1. "Average Prediction Accuracy versus Average SPS Score for alignments of 200 and 1000 sequences from 238 Pfam families" (Le et al. 2017) / CC BY 4.0

The tree building was done in MEGA-X for the reasons of the program being used in the exercises during the course and thus being somewhat familiar program. Tree method was chosen to be ML tree for the reasons described earlier. The ML seems to be the best non Bayesian approach to construct a robust phylogenetic tree and a bootstrap method was chosen for the test of phylogeny based on the hypothesis that the tree quality could be more solid after bootstrapping (Douady et al. 2003; Holder & Lewis 2003).

The substitution models come in many and to the author it was (and is) unclear what is the best method to be used. During the course there was a mention that the most common could be the best.

Clear is, that inappropriate substitution model can create a poor phylogenetic tree (Arenas 2015) but how good are the K2P and WAG and are they considered to be cheating is unclear to the author. WAG seems to be a general model and work nicely, but so does JTT (Le et al. 2008). Author couldn't interpret the data in the article of Le et al. (2008) to decide is WAG better than JTT, but based on the article there are better models than those mentioned. More advanced models aren't however in MEGA (Arenas 2015). For cDNA the K2P is said to be the most used model but it might not be the most rigorous one (Nishimaki & Sato 2019).

Identifying the most relevant SNPs from the vast list of SNPs is not that simple, as there seem to be at least tens of mutations that have a significant affect to LPL. Most significant might be the ones that are listed to OMIM and that lead to the deletion and thereby deficiency of the LPL. A thing to consider is also the UEF's and BluePrint genetics' list of harmful mutations. However I wasn't able to decipher the UEF's codon based idea and unclear was also why BluePrint Genetics' mutations are categorized as benign by the Ensembl. For the rs328 Ensembl shows two literature references that categorize it to cause familial lipoprotein lipase deficiency, but for the rs540525285 there aren't any articles listed and OMIM doesn't have information of it.

Several possibilities for error exist starting from the very beginning of the work. The BLAST search might have gone wrong albeit default settings were used. If the scoring rules would have been changed, it might have altered the results. However there was a clear consistency among the results: all found genes participate to lipid metabolism and more precisely to catabolism of triglycerides.

A great amount of uncertainty was in the selection of the most important SNPs to the VEP analyze and interpretation of the VEP results is - in my opinion - too shallow. The huge amount of mutations did confuse to some degree.

The MSA part is again in extremely weak base: there is no certainty of the correctness of the used sequences and sequence pruning wasn't done properly. I did ask instructions of what sequences to include but I'm not certain if the amount of sequences was right or not. The main transcripts were used but whether or not their alignment is correct is a question to be asked. Other questions like should the gap handling been done or should some of the sequences been truncated can be risen as well.

In tree building the possible error sources are the quality of the MSA and the correctness of the used methods. It is possible that 5000 bootstraps is far from being enough and the cutoff value for the partial deletion might have inappropriate. I did some sketches of the trees with low bootstrap values and varying cutoff percentages but as they were just sketches there isn't certainty that the final selections (5000 bootstraps and 30% cutoff) were good enough.

When it comes to the substitution models, things become even more uncertain. It should be so that when one uses specific models one has a well understanding of how the models work/differ from each other. I did find original papers to the used substitution models and did find some papers that describe methods in general but paid too little attention to the description of the models. Particularly the effect of discrete Gamma distribution (+G) and evolutionarily invariability (+I) are still unclear to me. But this was only a try based on curiosity to find the differences.

This was an interesting work to do, but it feels that I spend too little effort to find out in depth knowledge about the methods. Yet I dare to say that this work taught me the meaning of valid MSA and gave me courage to "play" with MSA sequences. By this I mean that I learnt that the order of the MSA can be changed and some eyesight or other QC should be done to get more reliable results. Also the phylogenetic part was somewhat new and after this work it feels that I have the possibility for further advance.

References:

Arenas M. Trends in substitution models of molecular evolution. Front. Genet. [Internet]. 2015 [Internet 2020 Nov 2];6. DOI:10.3389/fgene.2015.00319

Blueprint Genetics, LPL single gene test. https://blueprintgenetics.com/tests/single-gene-tests/lpl-single-gene-test-2/

Birrane G., Beigneux A.P., Dwyer B. et al. Structure of the lipoprotein lipase–GPIHBP1 complex that mediates plasma triglyceride hydrolysis. Proc. Natl. Acad. Sci. 2019;116(5):1723–32. DOI:10.1073/pnas.1817984116

Douady C.J., Delsuc F., Boucher Y. et al. Comparison of Bayesian and Maximum Likelihood Bootstrap Measures of Phylogenetic Reliability. Mol. Biol. Evol. Oxford Academic; 2003;20(2):248–54. DOI:10.1093/molbev/msg042

Hall B.G. Building Phylogenetic Trees from Molecular Data with MEGA. Mol. Biol. Evol. Oxford Academic; 2013;30(5):1229–35. DOI:10.1093/molbev/mst012

Holder M. & Lewis P.O. Phylogeny estimation: traditional and Bayesian approaches. Nat. Rev. Genet. Nature Publishing Group; 2003;4(4):275–84. DOI:10.1038/nrg1044

Horton J.D. Intravascular triglyceride lipolysis becomes crystal clear. Proc. Natl. Acad. Sci. 2019;116(5):1480–2. DOI:10.1073/pnas.1820330116

Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 1980;16(2):111–20. DOI:10.1007/BF01731581

Le Q., Sievers F. & Higgins D.G. Protein multiple sequence alignment benchmarking through secondary structure prediction. Bioinformatics. Oxford Academic; 2017;33(9):1331–7. DOI:10.1093/bioinformatics/btw840

Le S.Q., Lartillot N. & Gascuel O. Phylogenetic mixture models for proteins. Philos. Trans. R. Soc. B Biol. Sci. 2008;363(1512):3965–76. DOI:10.1098/rstb.2008.0180

Nishimaki T. & Sato K. An Extension of the Kimura Two-Parameter Model to the Natural Evolutionary Process. J. Mol. Evol. 2019;87(1):60–7. DOI:10.1007/s00239-018-9885-1

Sievers F. & Higgins D.G. Clustal Omega for making accurate alignments of many protein sequences. Protein Sci. 2018;27(1):135–45. DOI:10.1002/pro.3290

UEF **Genome Center of Eastern Finland,** Hypertriglyseridemiat. Retrieved 2.11.2020 from: https://sites.uef.fi/genediagnostics/monogeeniset-sairaudet/hypertriglyseridemiat/

Whelan S. & Goldman N. A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach. Mol. Biol. Evol. Oxford Academic; 2001;18(5):691–9. DOI:10.1093/oxfordjournals.molbev.a003851

Databases:

The databases used for this work are listed below as links to their main pages. Whenever a database was used as a reference, a specific database id was denoted (e.g. OMIM: 609708, ORPHA:444490, Ensembl xxyyyzz).

Alliance of Genome Resources (LPL entry): https://www.alliancegenome.org/gene/HGNC:6677#summary

Ensembl & Ensembl VEP: https://www.ensembl.org/index.html

EMBL-EBI, ClustalΩ: https://www.ebi.ac.uk/Tools/msa/clustalo/

GeneCards®: The Human Gene Database: https://www.genecards.org/

InterPro: https://www.ebi.ac.uk/interpro/

NCBI BLAST: https://blast.ncbi.nlm.nih.gov/Blast.cgi

NCBI Conserved Domains and Protein Classification database (CDD/SPARCLE): https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml

Online Mendelian Inheritance in Man, OMIM: https://www.omim.org/

Orphanet: https://www.orpha.net/consor/cgi-bin/index.php

UniProtKB: https://www.uniprot.org/

APPENDIX 1. THE BLASTP RESULTS FOR PARTIAL LPL

>NP_000228.1:28-475 lipoprotein lipase precursor [Homo sapiens]
ADQRRDFIDIESKFALRTPEDTAEDTCHLIPGVAESVATCHFNHSSKTFMVIHGWTVTGMYESWVPKLVAALYKREPDSN
VIVVDWLSRAQEHYPVSAGYTKLVGQDVARFINWMEEEFNYPLDNVHLLGYSLGAHAAGIAGSLTNKKVNRITGLDPAGP
NFEYAEAPSRLSPDDADFVDVLHTFTRGSPGRSIGIQKPVGHVDIYPNGGTFQPGCNIGEAIRVIAERGLGDVDQLVKCS
HERSIHLFIDSLLNEENPSKAYRCSSKEAFEKGLCLSCRKNRCNNLGYEINKVRAKRSSKMYLKTRSQMPYKVFHYQVKI
HFSGTESETHTNQAFEISLYGTVAESENIPFTLPEVSTNKTYSFLIYTEVDIGELLMLKLKWKSDSYFSWSDWWSSPGFA
IQKIRVKAGETQKKVIFCSREKVSHLQKGKAPAVFVKCHDKSLNKKSG
>NP_006024.1:50-485 endothelial lipase isoform 1 precursor [Homo sapiens]
RFNLRTSKDPEHEGCYLSVGHSQPLEDCSFNMTAKTFFIIHGWTMSGIFENWLHKLVSALHTREKDANVVVVDWLPLAHQ
LYTDAVNNTRVVGHSIARMLDWLQEKDDFSLGNVHLIGYSLGAHVAGYAGNFVKGTVGRITGLDPAGPMFEGADIHKRLS
PDDADFVDVLHTYTRSFGLSIGIQMPVGHIDIYPNGGDFQPGCGLNDVLGSIAYGTITEVVKCEHERAVHLFVDSLVNQD
KPSFAFQCTDSNRFKKGICLSCRKNRCNSIGYNAKKMRNKRNSKMYLKTRAGMPFRVYHYQMKIHVFSYKNMGEIEPTFY
VTLYGTNADSQTLPLEIVERIEQNATNTFLVYTEEDLGDLLKIQLTWEGASQSWYNLWKEFRSYLSQPRNPGRELNIRRI
RVKSGETQRKLTFCTEDPENTSISPGRELWFRKCRD
>XP_005258447.1:86-521 endothelial lipase isoform X1 [Homo sapiens]
RFNLRTSKDPEHEGCYLSVGHSQPLEDCSFNMTAKTFFIIHGWTMSGIFENWLHKLVSALHTREKDANVVVVDWLPLAHQ
LYTDAVNNTRVVGHSIARMLDWLQEKDDFSLGNVHLIGYSLGAHVAGYAGNFVKGTVGRITGLDPAGPMFEGADIHKRLS
PDDADFVDVLHTYTRSFGLSIGIQMPVGHIDIYPNGGDFQPGCGLNDVLGSIAYGTITEVVKCEHERAVHLFVDSLVNQD
KPSFAFQCTDSNRFKKGICLSCRKNRCNSIGYNAKKMRNKRNSKMYLKTRAGMPFRVYHYQMKIHVFSYKNMGEIEPTFY
VTLYGTNADSQTLPLEIVERIEQNATNTFLVYTEEDLGDLLKIQLTWEGASQSWYNLWKEFRSYLSQPRNPGRELNIRRI
RVKSGETQRKLTFCTEDPENTSISPGRELWFRKCRD
>XP_011524569.1:2-405 endothelial lipase isoform X3 [Homo sapiens]
TAKTFFIIHGWTMSGIFENWLHKLVSALHTREKDANVVVVDWLPLAHQLYTDAVNNTRVVGHSIARMLDWLQEKDDFSLG
NVHLIGYSLGAHVAGYAGNFVKGTVGRITGLDPAGPMFEGADIHKRLSPDDADFVDVLHTYTRSFGLSIGIQMPVGHIDI
YPNGGDFQPGCGLNDVLGSIAYGTITEVVKCEHERAVHLFVDSLVNQDKPSFAFQCTDSNRFKKGICLSCRKNRCNSIGY
NAKKMRNKRNSKMYLKTRAGMPFRVYHYQMKIHVFSYKNMGEIEPTFYVTLYGTNADSQTLPLEIVERIEQNATNTFLVY
TEEDLGDLLKIQLTWEGASQSWYNLWKEFRSYLSQPRNPGRELNIRRIRVKSGETQRKLTFCTEDPENTSISPGRELWFR
KCRD
>XP_006720565.1:15-448 hepatic triacylglycerol lipase isoform X3 [Homo sapiens]
CQIRINHPDTLQECGFNSSLPLVMIIHGWSVDGVLENWIWQMVAALKSQPAQPVNVGLVDWITLAHDHYTIAVRNTRLVG
KEVAALLRWLEESVQLSRSHVHLIGYSLGAHVSGFAGSSIGGTHKIGRITGLDAAGPLFEGSAPSNRLSPDDANFVDAIH
TFTREHMGLSVGIKQPIGHYDFYPNGGSFQPGCHFLELYRHIAQHGFNAITQTIKCSHERSVHLFIDSLLHAGTQSMAYP
CGDMNSFSQGLCLSCKKGRCNTLGYHVRQEPRSKSKRLFLVTRAQSPFKVYHYQFKIQFINQTETPIQTTFTMSLLGTKE
KMQKIPITLGKGIASNKTYSFLITLDVDIGELIMIKFKWENSAVWANVWDTVQTIIPWSTGPRHSGLVLKTIRVKAGETQ
QRMTFCSENTDDLLLRPTQEKIFVKCEIKSKTSK
>NP_000227.2:62-495 hepatic triacylglycerol lipase precursor [Homo sapiens]
CQIRINHPDTLQECGFNSSLPLVMIIHGWSVDGVLENWIWQMVAALKSQPAQPVNVGLVDWITLAHDHYTIAVRNTRLVG
KEVAALLRWLEESVQLSRSHVHLIGYSLGAHVSGFAGSSIGGTHKIGRITGLDAAGPLFEGSAPSNRLSPDDANFVDAIH
TFTREHMGLSVGIKQPIGHYDFYPNGGSFQPGCHFLELYRHIAQHGFNAITQTIKCSHERSVHLFIDSLLHAGTQSMAYP
CGDMNSFSQGLCLSCKKGRCNTLGYHVRQEPRSKSKRLFLVTRAQSPFKVYHYQFKIQFINQTETPIQTTFTMSLLGTKE
KMQKIPITLGKGIASNKTYSFLITLDVDIGELIMIKFKWENSAVWANVWDTVQTIIPWSTGPRHSGLVLKTIRVKAGETQ
QRMTFCSENTDDLLLRPTQEKIFVKCEIKSKTSK
>XP_005254431.2:74-507 hepatic triacylglycerol lipase isoform X1 [Homo sapiens]
CQIRINHPDTLQECGFNSSLPLVMIIHGWSVDGVLENWIWQMVAALKSQPAQPVNVGLVDWITLAHDHYTIAVRNTRLVG
KEVAALLRWLEESVQLSRSHVHLIGYSLGAHVSGFAGSSIGGTHKIGRITGLDAAGPLFEGSAPSNRLSPDDANFVDAIH
TFTREHMGLSVGIKQPIGHYDFYPNGGSFQPGCHFLELYRHIAQHGFNAITQTIKCSHERSVHLFIDSLLHAGTQSMAYP
CGDMNSFSQGLCLSCKKGRCNTLGYHVRQEPRSKSKRLFLVTRAQSPFKVYHYQFKIQFINQTETPIQTTFTMSLLGTKE
KMQKIPITLGKGIASNKTYSFLITLDVDIGELIMIKFKWENSAVWANVWDTVQTIIPWSTGPRHSGLVLKTIRVKAGETQ
QRMTFCSENTDDLLLRPTQEKIFVKCEIKSKTSK
>XP_016877665.1:74-367 hepatic triacylglycerol lipase isoform X4 [Homo sapiens]
CQIRINHPDTLQECGFNSSLPLVMIIHGWSVDGVLENWIWQMVAALKSQPAQPVNVGLVDWITLAHDHYTIAVRNTRLVG
KEVAALLRWLEESVQLSRSHVHLIGYSLGAHVSGFAGSSIGGTHKIGRITGLDAAGPLFEGSAPSNRLSPDDANFVDAIH
TFTREHMGLSVGIKQPIGHYDFYPNGGSFQPGCHFLELYRHIAQHGFNAITQTIKCSHERSVHLFIDSLLHAGTQSMAYP
CGDMNSFSQGLCLSCKKGRCNTLGYHVRQEPRSKSKRLFLVTRAQSPFKGFQLE
>XP_016881584.1:2-288 endothelial lipase isoform X4 [Homo sapiens]
FEGADIHKRLSPDDADFVDVLHTYTRSFGLSIGIQMPVGHIDIYPNGGDFQPGCGLNDVLGSIAYGTITEVVKCEHERAV
HLFVDSLVNQDKPSFAFQCTDSNRFKKGICLSCRKNRCNSIGYNAKKMRNKRNSKMYLKTRAGMPFRVYHYQMKIHVFSY
KNMGEIEPTFYVTLYGTNADSQTLPLEIVERIEQNATNTFLVYTEEDLGDLLKIQLTWEGASQSWYNLWKEFRSYLSQPR
NPGRELNIRRIRVKSGETQRKLTFCTEDPENTSISPGRELWFRKCRD
>XP_011524567.1:86-447 endothelial lipase isoform X2 [Homo sapiens]
RFNLRTSKDPEHEGCYLSVGHSQPLEDCSFNMTAKTFFIIHGWTMSGIFENWLHKLVSALHTREKDANVVVVDWLPLAHQ
LYTDAVNNTRVVGHSIARMLDWLQEKDDFSLGNVHLIGYSLGAHVAGYAGNFVKGTVGRITAITEVVKCEHERAVHLFVD
SLVNQDKPSFAFQCTDSNRFKKGICLSCRKNRCNSIGYNAKKMRNKRNSKMYLKTRAGMPFRVYHYQMKIHVFSYKNMGE
IEPTFYVTLYGTNADSQTLPLEIVERIEQNATNTFLVYTEEDLGDLLKIQLTWEGASQSWYNLWKEFRSYLSQPRNPGRE
LNIRRIRVKSGETQRKLTFCTEDPENTSISPGRELWFRKCRD

>NP_001294935.1:50-411 endothelial lipase isoform 2 precursor [Homo sapiens]
RFNLRTSKDPEHEGCYLSVGHSQPLEDCSFNMTAKTFFIIHGWTMSGIFENWLHKLVSALHTREKDANVVVVDWLPLAHQ
LYTDAVNNTRVVGHSIARMLDWLQEKDDFSLGNVHLIGYSLGAHVAGYAGNFVKGTVGRITAITEVVKCEHERAVHLFVD
SLVNQDKPSFAFQCTDSNRFKKGICLSCRKNRCNSIGYNAKKMRNKRNSKMYLKTRAGMPFRVYHYQMKIHVFSYKNMGE
IEPTFYVTLYGTNADSQTLPLEIVERIEQNATNTFLVYTEEDLGDLLKIQLTWEGASQSWYNLWKEFRSYLSQPRNPGRE
LNIRRIRVKSGETQRKLTFCTEDPENTSISPGRELWFRKCRD
>NP_000927.1:50-454 pancreatic triacylglycerol lipase precursor [Homo sapiens]
DVNTRFLLYTNENPNNFQEVAADSSSISGSNFKTNRKTRFIIHGFIDKGEENWLANVCKNLFKVESVNCICVDWKGGSRT
GYTQASQNIRIVGAEVAYFVEFLQSAFGYSPSNVHVIGHSLGAHAAGEAGRRTNGTIGRITGLDPAEPCFQGTPELVRLD
PSDAKFVDVIHTDGAPIVPNLGFGMSQVVGHLDFFPNGGVEMPGCKKNILSQIVDIDGIWEGTRDFAACNHLRSYKYYTD
SIVNPDGFAGFPCASYNVFTANKCFPCPSGGCPQMGHYADRYPGKTNDVGQKFYLDTGDASNFARWRYKVSVTLSGKKVT
GHILVSLFGNKGNSKQYEIFKGTLKPDSTHSNEFDSDVDVGDLQMVKFIWYNNVINPTLPRVGASKIIVETNVGKQFNFC
SPETV
>NP_001290064.1:52-456 inactive pancreatic lipase-related protein 1 precursor [Homo sapiens]
IGTRFLLYTNENPNNFQILLLSDPSTIEASNFQMDRKTRFIIHGFIDKGDESWVTDMCKKLFEVEEVNCICVDWKKGSQA
TYTQAANNVRVVGAQVAQMLDILLTEYSYPPSKVHLIGHSLGAHVAGEAGSKTPGLSRITGLDPVEASFESTPEEVRLDP
SDADFVDVIHTDAAPLIPFLGFGTNQQMGHLDFFPNGGESMPGCKKNALSQIVDLDGIWAGTRDFVACNHLRSYKYYLES
ILNPDGFAAYPCTSYKSFESDKCFPCPDQGCPQMGHYADKFAGRTSEEQQKFFLNTGEASNFARWRYGVSITLSGRTATG
QIKVALFGNKGNTHQYSIFRGILKPGSTHSYEFDAKLDVGTIEKVKFLWNNNVINPTLPKVGATKITVQKGEEKTVYNFC
SEDTV
>NP_005387.3:51-458 pancreatic lipase-related protein 2 precursor [Homo sapiens]
DIDTRFLLYTNENPNNFQLITGTEPDTIEASNFQLDRKTRFIIHGFLDKAEDSWPSDMCKKMFEVEKVNCICVDWRHGSR
AMYTQAVQNIRVVGAETAFLIQALSTQLGYSLEDVHVIGHSLGAHTAAEAGRRLGGRVGRITGLDPAGPCFQDEPEEVRL
DPSDAVFVDVIHTDSSPIVPSLGFGMSQKVGHLDFFPNGGKEMPGCKKNVLSTITDIDGIWEGIGGFVSCNHLRSFEYYS
SSVLNPDGFLGYPCASYDEFQESKCFPCPAEGCPKMGHYADQFKGKTSAVEQTFFLNTGESGNFTSWRYKVSVTLSGKEK
VNGYIRIALYGSNENSKQYEIFKGSLKPDASHTCAIDVDFNVGKIQKVKFLWNKRGINLSEPKLGASQITVQSGEDGTEY
NFCSSDTV
>NP_001011709.2:75-454 pancreatic lipase-related protein 3 precursor [Homo sapiens]
SSTIQASYFGTDKITRINIAGWKTDGKWQRDMCNVLLQLEDINCINLDWINGSREYIHAVNNLRVVGAEVAYFIDVLMKK
FEYSPSKVHLIGHSLGAHLAGEAGSRIPGLGRITGLDPAGPFFHNTPKEVRLDPSDANFVDVIHTNAARILFELGVGTID
ACGHLDFYPNGGKHMPGCEDLITPLLKFNFNAYKKEMASFFDCNHARSYQFYAESILNPDAFIAYPCRSYTSFKAGNCFF
CSKEGCPTMGHFADRFHFKNMKTNGSHYFLNTGSLSPFARWRHKLSVKLSGSEVTQGTVFLRVGGAVRKTGEFAIVSGKL
EPGMTYTKLIDADVNVGNITSVQFIWKKHLFEDSQNKLGAEMVINTSGKYGYKSTFCSQD
>XP_011510832.1:16-292 lipase member H isoform X3 [Homo sapiens]
SSAFGNLNVTKKTTFIVHGFRPTGSPPVWMDDLVKGLLSVEDMNVVVVDWNRGATTLIYTHASSKTRKVAMVLKEFIDQM
LAEGASLDDIYMIGVSLGAHISGFVGEMYDGWLGRITGLDPAGPLFNGKPHQDRLDPSDAQFVDVIHSDTDALGYKEPLG
NIDFYPNGGLDQPGCPKTILGGFQYFKCDHQRSVYLYLSSLRESCTITAYPCDSYQDYRNGKCVSCGTSQKESCPLLGYY
ADNWKDHLRGKDPPMTKAFFDTAEESPFCMYHYFVDI
>NP_640341.1:59-335 lipase member H precursor [Homo sapiens]
SSAFGNLNVTKKTTFIVHGFRPTGSPPVWMDDLVKGLLSVEDMNVVVVDWNRGATTLIYTHASSKTRKVAMVLKEFIDQM
LAEGASLDDIYMIGVSLGAHISGFVGEMYDGWLGRITGLDPAGPLFNGKPHQDRLDPSDAQFVDVIHSDTDALGYKEPLG
NIDFYPNGGLDQPGCPKTILGGFQYFKCDHQRSVYLYLSSLRESCTITAYPCDSYQDYRNGKCVSCGTSQKESCPLLGYY
ADNWKDHLRGKDPPMTKAFFDTAEESPFCMYHYFVDI
>XP_011537580.1:10-349 pancreatic lipase-related protein 3 isoform X3 [Homo sapiens]
DINCINLDWINGSREYIHAVNNLRVVGAEVAYFIDVLMKKFEYSPSKVHLIGHSLGAHLAGEAGSRIPGLGRITGLDPAG
PFFHNTPKEVRLDPSDANFVDVIHTNAARILFELGVGTIDACGHLDFYPNGGKHMPGCEDLITPLLKFNFNAYKKEMASF
FDCNHARSYQFYAESILNPDAFIAYPCRSYTSFKAGNCFFCSKEGCPTMGHFADRFHFKNMKTNGSHYFLNTGSLSPFAR
WRHKLSVKLSGSEVTQGTVFLRVGGAVRKTGEFAIVSGKLEPGMTYTKLIDADVNVGNITSVQFIWKKHLFEDSQNKLGA
EMVINTSGKYGYKSTFCSQD
>NP_001289930.1:69-361 lipase member I isoform 4 [Homo sapiens]
NFNTQKKTVWLIHGYRPVGSIPLWLQNFVRILLNEEDMNVIVVDWSRGATTFIYNRAVKNTRKVAVSLSVHIKNLLKHGA
SLDNFHFIGVSLGAHISGFVGKIFHGQLGRITGLDPAGPRFSRKPPYSRLDYTDAKFVDVIHSDSNGLGIQEPLGHIDFY
PNGGNKQPGCPKSIFSGIQFIKCNHQRAVHLFMASLETNCNFISFPCRSYKDYKTSLCVDCDCFKEKSCPRLGYQAKLFK
GVLKERMEGRPLRTTVFLDTSGTYPFCNHHFAGIILYLKTERKCFLIQTHVHQ
>XP_011537579.1:16-355 pancreatic lipase-related protein 3 isoform X2 [Homo sapiens]
DINCINLDWINGSREYIHAVNNLRVVGAEVAYFIDVLMKKFEYSPSKVHLIGHSLGAHLAGEAGSRIPGLGRITGLDPAG
PFFHNTPKEVRLDPSDANFVDVIHTNAARILFELGVGTIDACGHLDFYPNGGKHMPGCEDLITPLLKFNFNAYKKEMASF
FDCNHARSYQFYAESILNPDAFIAYPCRSYTSFKAGNCFFCSKEGCPTMGHFADRFHFKNMKTNGSHYFLNTGSLSPFAR
WRHKLSVKLSGSEVTQGTVFLRVGGAVRKTGEFAIVSGKLEPGMTYTKLIDADVNVGNITSVQFIWKKHLFEDSQNKLGA
EMVINTSGKYGYKSTFCSQD
>NP_945347.3:24-298 lipase member I isoform 5 [Homo sapiens]
NFNTQKKTVWLIHGYRPVGSIPLWLQNFVRILLNEEDMNVIVVDWSRGATTFIYNRAVKNTRKVAVSLSVHIKNLLKHGA
SLDNFHFIGVSLGAHISGFVGKIFHGQLGRITGLDPAGPRFSRKPPYSRLDYTDAKFVDVIHSDSNGLGIQEPLGHIDFY
PNGGNKQPGCPKSIFSGIQFIKCNHQRAVHLFMASLETNCNFISFPCRSYKDYKTSLCVDCDCFKEKSCPRLGYQAKLFK
GVLKERMEGRPLRTTVFLDTSGTYPFCTYYFVLSI
>XP_006724028.1:98-372 lipase member I isoform X1 [Homo sapiens]
NFNTQKKTVWLIHGYRPVGSIPLWLQNFVRILLNEEDMNVIVVDWSRGATTFIYNRAVKNTRKVAVSLSVHIKNLLKHGA
SLDNFHFIGVSLGAHISGFVGKIFHGQLGRITGLDPAGPRFSRKPPYSRLDYTDAKFVDVIHSDSNGLGIQEPLGHIDFY
PNGGNKQPGCPKSIFSGIQFIKCNHQRAVHLFMASLETNCNFISFPCRSYKDYKTSLCVDCDCFKEKSCPRLGYQAKLFK
GVLKERMEGRPLRTTVFLDTSGTYPFCTYYFVLSI

>NP_001289927.1:69-343 lipase member I isoform 1 precursor [Homo sapiens]
NFNTQKKTVWLIHGYRPVGSIPLWLQNFVRILLNEEDMNVIVVDWSRGATTFIYNRAVKNTRKVAVSLSVHIKNLLKHGA
SLDNFHFIGVSLGAHISGFVGKIFHGQLGRITGLDPAGPRFSRKPPYSRLDYTDAKFVDVIHSDSNGLGIQEPLGHIDFY
PNGGNKQPGCPKSIFSGIQFIKCNHQRAVHLFMASLETNCNFISFPCRSYKDYKTSLCVDCDCFKEKSCPRLGYQAKLFK
GVLKERMEGRPLRTTVFLDTSGTYPFCTYYFVLSI
>NP_001289929.1:69-315 lipase member I isoform 3 precursor [Homo sapiens]
NFNTQKKTVWLIHGYRPVGSIPLWLQNFVRILLNEEDMNVIVVDWSRGATTFIYNRAVKNTRKVAVSLSVHIKNLLKHGA
SLDNFHFIGVSLGAHISGFVGKIFHGQLGRITGLDPAGPRFSRKPPYSRLDYTDAKFVDVIHSDSNGLGIQEPLGHIDFY
PNGGNKQPGCPKSIFSGIQFIKCNHQRAVHLFMASLETNCNFISFPCRSYKDYKTSLCVDCDCFKEKSCPRLGYQAKLFK
GVLKERM
>XP_011538171.1:1-307 inactive pancreatic lipase-related protein 1 isoform X1 [Homo sapiens]
MLDILLTEYSYPPSKVHLIGHSLGAHVAGEAGSKTPGLSRITGLDPVEASFESTPEEVRLDPSDADFVDVIHTDAAPLIP
FLGFGTNQQMGHLDFFPNGGESMPGCKKNALSQIVDLDGIWAGTRDFVACNHLRSYKYYLESILNPDGFAAYPCTSYKSF
ESDKCFPCPDQGCPQMGHYADKFAGRTSEEQQKFFLNTGEASNFARWRYGVSITLSGRTATGQIKVALFGNKGNTHQYSI
FRGILKPGSTHSYEFDAKLDVGTIEKVKFLWNNNVINPTLPKVGATKITVQKGEEKTVYNFCSEDTV
>NP_001280154.1:32-331 phospholipase A1 member A isoform 4 [Homo sapiens]
DLKVQFLLFVPSNPSCGQLVEGSSDLQNSGFNATLGTKLIIHGFRVLGTKPSWIDTFIRTLLRATNANVIAVDWIYGSTG
VYFSAVKNVIKLSLEISLFLNKLLVLGVSESSIHIIGVSLGAHVGGMVGQLFGGQLGQITGLDPAGPEYTRASVEERLDA
GDALFVEAIHTDTDNLGIRIPVGHVDYFVNGGQDQPGCPTFFYAGYSYLICDHMRAVHLYISALENSCPLMAFPCASYKA
FLAGRCLDCFNPFLLSCPRIGLVEQGGVKIEPLPKEVKVYLLTTSSAPYCMHHSLVEFHL
>NP_056984.1:48-347 phospholipase A1 member A isoform 1 precursor [Homo sapiens]
DLKVQFLLFVPSNPSCGQLVEGSSDLQNSGFNATLGTKLIIHGFRVLGTKPSWIDTFIRTLLRATNANVIAVDWIYGSTG
VYFSAVKNVIKLSLEISLFLNKLLVLGVSESSIHIIGVSLGAHVGGMVGQLFGGQLGQITGLDPAGPEYTRASVEERLDA
GDALFVEAIHTDTDNLGIRIPVGHVDYFVNGGQDQPGCPTFFYAGYSYLICDHMRAVHLYISALENSCPLMAFPCASYKA
FLAGRCLDCFNPFLLSCPRIGLVEQGGVKIEPLPKEVKVYLLTTSSAPYCMHHSLVEFHL
>NP_001193889.1:48-331 phospholipase A1 member A isoform 2 precursor [Homo sapiens]
DLKVQFLLFVPSNPSCGQLVEGSSDLQNSGFNATLGTKLIIHGFRVLGTKPSWIDTFIRTLLRATNANVIAVDWIYGSTG
VYFSAVKNVLGVSESSIHIIGVSLGAHVGGMVGQLFGGQLGQITGLDPAGPEYTRASVEERLDAGDALFVEAIHTDTDNL
GIRIPVGHVDYFVNGGQDQPGCPTFFYAGYSYLICDHMRAVHLYISALENSCPLMAFPCASYKAFLAGRCLDCFNPFLLS
CPRIGLVEQGGVKIEPLPKEVKVYLLTTSSAPYCMHHSLVEFHL
>NP_001366494.1:69-238 lipase member I isoform 6 [Homo sapiens]
NFNTQKKTVWLIHGYRPVGSIPLWLQNFVRILLNEEDMNVIVVDWSRGATTFIYNRAVKNTRKVAVSLSVHIKNLLKHGA
SLDNFHFIGVSLGAHISGFVGKIFHGQLGRITGLDPAGPRFSRKPPYSRLDYTDAKFVDVIHSDSNGLGIQEPLGHIDFY
PNGGNKQPGC
>XP_006713592.1:59-305 lipase member H isoform X1 [Homo sapiens]
SSAFGNLNVTKKTTFIVHGFRPTGSPPVWMDDLVKGLLSVEDMNVVVVDWNRGATTLIYTHASSKTRKVAMVLKEFIDQM
LAEGASLDDIYMIGVSLGAHISGFVGEMYDGWLGRITGLDPAGPLFNGKPHQDRLDPSDAQFVDVIHSDTDGFQYFKCDH
QRSVYLYLSSLRESCTITAYPCDSYQDYRNGKCVSCGTSQKESCPLLGYYADNWKDHLRGKDPPMTKAFFDTAEESPFCM
YHYFVDI
>XP_011537578.1:75-252 pancreatic lipase-related protein 3 isoform X1 [Homo sapiens]
SSTIQASYFGTDKITRINIAGWKTDGKWQRDMCNVLLQLEDINCINLDWINGSREYIHAVNNLRVVGAEVAYFIDVLMKK
FEYSPSKVHLIGHSLGAHLAGEAGSRIPGLGRITGLDPAGPFFHNTPKEVRLDPSDANFVDVIHTNAARILFELGVGTID
ACGHLDFYPNGGKHMPGC
>XP_016861341.1:59-301 lipase member H isoform X2 [Homo sapiens]
SSAFGNLNVTKKTTFIVHGFRPTGSPPVWMDDLVKGLLSVEDMNVVVVDWNRGATTLIYTHASSKTRKVAMVLKEFIDQM
LAEGASLDDIYMIGVSLGAHISGFVGEMYDGWLGRITALGYKEPLGNIDFYPNGGLDQPGCPKTILGGFQYFKCDHQRSV
YLYLSSLRESCTITAYPCDSYQDYRNGKCVSCGTSQKESCPLLGYYADNWKDHLRGKDPPMTKAFFDTAEESPFCMYHYF
VDI
>NP_001289928.1:69-313 lipase member I isoform 2 [Homo sapiens]
NFNTQKKTVWLIHGYRPVGSIPLWLQNFVRILLNEEDMNVIVVDWSRGATTFIYNRAVKNTRKVAVSLSVHIKNLLKHGA
SLDNFHFIGVSLGAHISGFVGKIFHGQLGRITGLDPAGPRFSRKPPYSRLDYTDAKFVDVIHSDSNGIQFIKCNHQRAVH
LFMASLETNCNFISFPCRSYKDYKTSLCVDCDCFKEKSCPRLGYQAKLFKGVLKERMEGRPLRTTVFLDTSGTYPFCTYY
FVLSI
>NP_001366495.1:15-178 lipase member I isoform 7 precursor [Homo sapiens]
SGLDPAGPRFSRKPPYSRLDYTDAKFVDVIHSDSNGLGIQEPLGHIDFYPNGGNKQPGCPKSIFSGIQFIKCNHQRAVHL
FMASLETNCNFISFPCRSYKDYKTSLCVDCDCFKEKSCPRLGYQAKLFKGVLKERMEGRPLRTTVFLDTSGTYPFCTYYF
VLSI
>NP_001193890.1:1-174 phospholipase A1 member A isoform 3 [Homo sapiens]
MVGQLFGGQLGQITGLDPAGPEYTRASVEERLDAGDALFVEAIHTDTDNLGIRIPVGHVDYFVNGGQDQPGCPTFFYAGY
SYLICDHMRAVHLYISALENSCPLMAFPCASYKAFLAGRCLDCFNPFLLSCPRIGLVEQGGVKIEPLPKEVKVYLLTTSS
APYCMHHSLVEFHL

APPENDIX 2. FIRST MSA WITH ALL BLASTED SEQUENCES.

```
CLUSTAL O(1.2.4) multiple sequence alignment


X5-LIPC                 -------------------------CQIRINHPDTLQECGFNSSLPLVMIIHGWSVDGV
LIPC                    -------------------------CQIRINHPDTLQECGFNSSLPLVMIIHGWSVDGV
X1-LIPC                 -------------------------CQIRINHPDTLQECGFNSSLPLVMIIHGWSVDGV
X6-LIPC                 -------------------------CQIRINHPDTLQECGFNSSLPLVMIIHGWSVDGV
LPL                     ADQRRDFIDIESKFALRTPEDTAEDTCHLIPGVAESVATCHFNHSSKTFMVIHGWTVTGM
LIPG-variant1           ------------RFNLRTSKDPEHEGCYLSVGHSQPLEDCSFNMTAKTFFIIHGWTMSGI
X1-LIPG                 ------------RFNLRTSKDPEHEGCYLSVGHSQPLEDCSFNMTAKTFFIIHGWTMSGI
X3-LIPG                 ---------------------------------------------TAKTFFIIHGWTMSGI
X4-LIPG                 ------------------------------------------------------------
X2-LIPG                 ------------RFNLRTSKDPEHEGCYLSVGHSQPLEDCSFNMTAKTFFIIHGWTMSGI
LIPG-variant2           ------------RFNLRTSKDPEHEGCYLSVGHSQPLEDCSFNMTAKTFFIIHGWTMSGI
X1-PNLIPRP3             ----------------------------------SSTIQASYFGTDKITRINIAGWKTDG-
PNLIPRP3                ----------------------------------SSTIQASYFGTDKITRINIAGWKTDG-
X3-PNLIPRP3             ------------------------------------------------------------
X2-PNLIPRP3             ------------------------------------------------------------
PNLIPRP2                --------DIDTRFLLYTNENPNNFQLIT-GTEPDTIEASNFQLDRKTRFIIHGFLDKAE
PNLIP                   --------DVNTRFLLYTNENPNNFQEV--AADSSSISGSNFKTNRKTRFIIHGFIDKGE
PNLIPRP1-variant        ---------IGTRFLLYTNENPNNFQILL-LSDPSTIEASNFQMDRKTRFIIHGFIDKGD
X1-PNLIPRP1             ------------------------------------------------------------
PLA1A-variant4          --------DLKVQFLLFVPSNPSCGQLVE---GSSDLQNSGFNATLGTKLIIHGFRVLGT
PLA1A-variant1          --------DLKVQFLLFVPSNPSCGQLVE---GSSDLQNSGFNATLGTKLIIHGFRVLGT
PLA1A-variant2          --------DLKVQFLLFVPSNPSCGQLVE---GSSDLQNSGFNATLGTKLIIHGFRVLGT
PLA1A-variant3          ------------------------------------------------------------
X2-LIPH                 ----------------------------------SSAFGNLNVTKKTTFIVHGFRPTGS
X3-LIPH                 ----------------------------------SSAFGNLNVTKKTTFIVHGFRPTGS
LIPH                    ----------------------------------SSAFGNLNVTKKTTFIVHGFRPTGS
X1-LIPH                 ----------------------------------SSAFGNLNVTKKTTFIVHGFRPTGS
LIPI-variant7           ------------------------------------------------------------
LIPI-variant.deltaE8    ----------------------------------------NFNTQKKTVWLIHGYRPVGS
LIPI-variant.deltaE7    ----------------------------------------NFNTQKKTVWLIHGYRPVGS
LIPI-variant6           ----------------------------------------NFNTQKKTVWLIHGYRPVGS
LIPI-variant2           ----------------------------------------NFNTQKKTVWLIHGYRPVGS
X1-LIP1                 ----------------------------------------NFNTQKKTVWLIHGYRPVGS
LIPI-variant.fl         ----------------------------------------NFNTQKKTVWLIHGYRPVGS
LIPI-variant.deltaE5    ----------------------------------------NFNTQKKTVWLIHGYRPVGS


X5-LIPC                 LENWIWQMVAALKSQPAQPVNVGLVDWITLAHD-HYTIAVRNTRLVGKEVAALLRWLEES
LIPC                    LENWIWQMVAALKSQPAQPVNVGLVDWITLAHD-HYTIAVRNTRLVGKEVAALLRWLEES
X1-LIPC                 LENWIWQMVAALKSQPAQPVNVGLVDWITLAHD-HYTIAVRNTRLVGKEVAALLRWLEES
X6-LIPC                 LENWIWQMVAALKSQPAQPVNVGLVDWITLAHD-HYTIAVRNTRLVGKEVAALLRWLEES
LPL                     YESWVPKLVAALYKRE-PDSNVIVVDWLSRAQE-HYPVSAGYTKLVGQDVARFINWMEEE
LIPG-variant1           FENWLHKLVSALHTRE-KDANVVVVDWLPLAHQ-LYTDAVNNTRVVGHSIARMLDWLQEK
X1-LIPG                 FENWLHKLVSALHTRE-KDANVVVVDWLPLAHQ-LYTDAVNNTRVVGHSIARMLDWLQEK
X3-LIPG                 FENWLHKLVSALHTRE-KDANVVVVDWLPLAHQ-LYTDAVNNTRVVGHSIARMLDWLQEK
X4-LIPG                 ------------------------------------------------------------
X2-LIPG                 FENWLHKLVSALHTRE-KDANVVVVDWLPLAHQ-LYTDAVNNTRVVGHSIARMLDWLQEK
LIPG-variant2           FENWLHKLVSALHTRE-KDANVVVVDWLPLAHQ-LYTDAVNNTRVVGHSIARMLDWLQEK
X1-PNLIPRP3             --KWQRDMCNVLLQLE--DINCINLDWINGSR--EYIHAVNNLRVVGAEVAYFIDVLMKK
PNLIPRP3                --KWQRDMCNVLLQLE--DINCINLDWINGSR--EYIHAVNNLRVVGAEVAYFIDVLMKK
X3-PNLIPRP3             ------------------DINCINLDWINGSR--EYIHAVNNLRVVGAEVAYFIDVLMKK
X2-PNLIPRP3             ------------------DINCINLDWINGSR--EYIHAVNNLRVVGAEVAYFIDVLMKK
PNLIPRP2                -DSWPSDMCKKMFEVE--KVNCICVDWRHGSRA-MYTQAVQNIRVVGAETAFLIQALSTQ
PNLIP                   -ENWLANVCKNLFKVE--SVNCICVDWKGGSRT-GYTQASQNIRIVGAEVAYFVEFLQSA
PNLIPRP1-variant        -ESWVTDMCKKLFEVE--EVNCICVDWKKGSQA-TYTQAANNVRVVGAQVAQMLDILLTE
X1-PNLIPRP1             ------------------------------------------------MLDILLTE
PLA1A-variant4          KPSWIDTFIRTLLRAT--NANVIAVDWIYGSTG-VYFSAVKNVIKLSLEISLFLNKLL-V
PLA1A-variant1          KPSWIDTFIRTLLRAT--NANVIAVDWIYGSTG-VYFSAVKNVIKLSLEISLFLNKLL-V
PLA1A-variant2          KPSWIDTFIRTLLRAT--NANVIAVDWIYGSTG-VYFSAVKNV-----------------
PLA1A-variant3          ------------------------------------------------------------
X2-LIPH                 PPVWMDDLVKGLLSVE--DMNVVVVDWNRGATTLIYTHASSKTRKVAMVLKEFIDQML-A
X3-LIPH                 PPVWMDDLVKGLLSVE--DMNVVVVDWNRGATTLIYTHASSKTRKVAMVLKEFIDQML-A
LIPH                    PPVWMDDLVKGLLSVE--DMNVVVVDWNRGATTLIYTHASSKTRKVAMVLKEFIDQML-A
X1-LIPH                 PPVWMDDLVKGLLSVE--DMNVVVVDWNRGATTLIYTHASSKTRKVAMVLKEFIDQML-A
LIPI-variant7           ------------------------------------------------------------
LIPI-variant.deltaE8    IPLWLQNFVRILLNEE--DMNVIVVDWSRGATTFIYNRAVKNTRKVAVSLSVHIKNLL-K
LIPI-variant.deltaE7    IPLWLQNFVRILLNEE--DMNVIVVDWSRGATTFIYNRAVKNTRKVAVSLSVHIKNLL-K
LIPI-variant6           IPLWLQNFVRILLNEE--DMNVIVVDWSRGATTFIYNRAVKNTRKVAVSLSVHIKNLL-K
LIPI-variant2           IPLWLQNFVRILLNEE--DMNVIVVDWSRGATTFIYNRAVKNTRKVAVSLSVHIKNLL-K
X1-LIP1                 IPLWLQNFVRILLNEE--DMNVIVVDWSRGATTFIYNRAVKNTRKVAVSLSVHIKNLL-K
LIPI-variant.fl         IPLWLQNFVRILLNEE--DMNVIVVDWSRGATTFIYNRAVKNTRKVAVSLSVHIKNLL-K
LIPI-variant.deltaE5    IPLWLQNFVRILLNEE--DMNVIVVDWSRGATTFIYNRAVKNTRKVAVSLSVHIKNLL-K


X5-LIPC                 VQLSRSHVHLIGYSLGAHVSGFAGSSIGGTHKIGRITGLDAAGPLFEGSAPSNRLSPDDA
LIPC                    VQLSRSHVHLIGYSLGAHVSGFAGSSIGGTHKIGRITGLDAAGPLFEGSAPSNRLSPDDA
X1-LIPC                 VQLSRSHVHLIGYSLGAHVSGFAGSSIGGTHKIGRITGLDAAGPLFEGSAPSNRLSPDDA
X6-LIPC                 VQLSRSHVHLIGYSLGAHVSGFAGSSIGGTHKIGRITGLDAAGPLFEGSAPSNRLSPDDA
LPL                     FNYPLDNVHLLGYSLGAHAAGIAGSLTNK--KVNRITGLDPAGPNFEYAEAPSRLSPDDA
LIPG-variant1           DDFSLGNVHLIGYSLGAHVAGYAGNFVKG--TVGRITGLDPAGPMFEGADIHKRLSPDDA
X1-LIPG                 DDFSLGNVHLIGYSLGAHVAGYAGNFVKG--TVGRITGLDPAGPMFEGADIHKRLSPDDA
```

```
X3-LIPG                 DDFSLGNVHLIGYSLGAHVAGYAGNFVKG--TVGRITGLDPAGPMFEGADIHKRLSPDDA
X4-LIPG                 ------------------------------------------FEGADIHKRLSPDDA
X2-LIPG                 DDFSLGNVHLIGYSLGAHVAGYAGNFVKG--TVGRITAI----------------------
LIPG-variant2           DDFSLGNVHLIGYSLGAHVAGYAGNFVKG--TVGRITAI----------------------
X1-PNLIPRP3             FEYSPSKVHLIGHSLGAHLAGEAGSRIPG---LGRITGLDPAGPFFHNTPKEVRLDPSDA
PNLIPRP3                FEYSPSKVHLIGHSLGAHLAGEAGSRIPG---LGRITGLDPAGPFFHNTPKEVRLDPSDA
X3-PNLIPRP3             FEYSPSKVHLIGHSLGAHLAGEAGSRIPG---LGRITGLDPAGPFFHNTPKEVRLDPSDA
X2-PNLIPRP3             FEYSPSKVHLIGHSLGAHLAGEAGSRIPG---LGRITGLDPAGPFFHNTPKEVRLDPSDA
PNLIPRP2                LGYSLEDVHVIGHSLGAHTAAEAGRRLGG--RVGRITGLDPAGPCFQDEPEEVRLDPSDA
PNLIP                   FGYSPSNVHVIGHSLGAHAAGEAGRRTNG--TIGRITGLDPAEPCFQGTPELVRLDPSDA
PNLIPRP1-variant        YSYPPSKVHLIGHSLGAHVAGEAGSKTPG---LSRITGLDPVEASFESTPEEVRLDPSDA
X1-PNLIPRP1             YSYPPSKVHLIGHSLGAHVAGEAGSKTPG---LSRITGLDPVEASFESTPEEVRLDPSDA
PLA1A-variant4          LGVSESSIHIIGVSLGAHVGGMVGQLFGG--QLGQITGLDPAGPEYTRASVEERLDAGDA
PLA1A-variant1          LGVSESSIHIIGVSLGAHVGGMVGQLFGG--QLGQITGLDPAGPEYTRASVEERLDAGDA
PLA1A-variant2          LGVSESSIHIIGVSLGAHVGGMVGQLFGG--QLGQITGLDPAGPEYTRASVEERLDAGDA
PLA1A-variant3          ----------------------MVGQLFGG--QLGQITGLDPAGPEYTRASVEERLDAGDA
X2-LIPH                 EGASLDDIYMIGVSLGAHISGFVGEMYDG--WLGRITA----------------------
X3-LIPH                 EGASLDDIYMIGVSLGAHISGFVGEMYDG--WLGRITGLDPAGPLFNGKPHQDRLDPSDA
LIPH                    EGASLDDIYMIGVSLGAHISGFVGEMYDG--WLGRITGLDPAGPLFNGKPHQDRLDPSDA
X1-LIPH                 EGASLDDIYMIGVSLGAHISGFVGEMYDG--WLGRITGLDPAGPLFNGKPHQDRLDPSDA
LIPI-variant7           ------------------------------------SGLDPAGPRFSRKPPYSRLDYTDA
LIPI-variant.deltaE8    HGASLDNFHFIGVSLGAHISGFVGKIFHG--QLGRITGLDPAGPRFSRKPPYSRLDYTDA
LIPI-variant.deltaE7    HGASLDNFHFIGVSLGAHISGFVGKIFHG--QLGRITGLDPAGPRFSRKPPYSRLDYTDA
LIPI-variant6           HGASLDNFHFIGVSLGAHISGFVGKIFHG--QLGRITGLDPAGPRFSRKPPYSRLDYTDA
LIPI-variant2           HGASLDNFHFIGVSLGAHISGFVGKIFHG--QLGRITGLDPAGPRFSRKPPYSRLDYTDA
X1-LIP1                 HGASLDNFHFIGVSLGAHISGFVGKIFHG--QLGRITGLDPAGPRFSRKPPYSRLDYTDA
LIPI-variant.fl         HGASLDNFHFIGVSLGAHISGFVGKIFHG--QLGRITGLDPAGPRFSRKPPYSRLDYTDA
LIPI-variant.deltaE5    HGASLDNFHFIGVSLGAHISGFVGKIFHG--QLGRITGLDPAGPRFSRKPPYSRLDYTDA

X5-LIPC                 NFVDAIHTFTR-EHMGLSVGIKQPIGHYDFYPNGGSFQPGCHFLELYRH--IAQHGFNAI
LIPC                    NFVDAIHTFTR-EHMGLSVGIKQPIGHYDFYPNGGSFQPGCHFLELYRH--IAQHGFNAI
X1-LIPC                 NFVDAIHTFTR-EHMGLSVGIKQPIGHYDFYPNGGSFQPGCHFLELYRH--IAQHGFNAI
X6-LIPC                 NFVDAIHTFTR-EHMGLSVGIKQPIGHYDFYPNGGSFQPGCHFLELYRH--IAQHGFNAI
LPL                     DFVDVLHTFTR-GSPGRSIGIQKPVGHVDIYPNGGTFQPGCNIGEAIRV--IAERGLGDV
LIPG-variant1           DFVDVLHTYTR--SFGLSIGIQMPVGHIDIYPNGGDFQPGCGLNDVLGS--I---AYGTI
X1-LIPG                 DFVDVLHTYTR--SFGLSIGIQMPVGHIDIYPNGGDFQPGCGLNDVLGS--I---AYGTI
X3-LIPG                 DFVDVLHTYTR--SFGLSIGIQMPVGHIDIYPNGGDFQPGCGLNDVLGS--I---AYGTI
X4-LIPG                 DFVDVLHTYTR--SFGLSIGIQMPVGHIDIYPNGGDFQPGCGLNDVLGS--I---AYGTI
X2-LIPG                 ------------------------------------------------------------
LIPG-variant2           ------------------------------------------------------------
X1-PNLIPRP3             NFVDVIHTNAARILFELGVGTIDACGHLDFYPNGGKHMPGC--------------------
PNLIPRP3                NFVDVIHTNAARILFELGVGTIDACGHLDFYPNGGKHMPGCEDLITPLLKFNFNAYKKEM
X3-PNLIPRP3             NFVDVIHTNAARILFELGVGTIDACGHLDFYPNGGKHMPGCEDLITPLLKFNFNAYKKEM
X2-PNLIPRP3             NFVDVIHTNAARILFELGVGTIDACGHLDFYPNGGKHMPGCEDLITPLLKFNFNAYKKEM
PNLIPRP2                VFVDVIHTDSSPIVPSLGFGMSQKVGHLDFFPNGGKEMPGCKKNVLSTI-TDIDGIWEGI
PNLIP                   KFVDVIHTDGAPIVPNLGFGMSQVVGHLDFFPNGGVEMPGCKKNILSQI-VDIDGIWEGT
PNLIPRP1-variant        DFVDVIHTDAAPLIPFLGFGTNQQMGHLDFFPNGGESMPGCKKNALSQI-VDLDGIWAGT
X1-PNLIPRP1             DFVDVIHTDAAPLIPFLGFGTNQQMGHLDFFPNGGESMPGCKKNALSQI-VDLDGIWAGT
PLA1A-variant4          LFVEAIHTDTD------NLGIRIPVGHVDYFVNGGQDQPGCPTFFYAGY-----------
PLA1A-variant1          LFVEAIHTDTD------NLGIRIPVGHVDYFVNGGQDQPGCPTFFYAGY-----------
PLA1A-variant2          LFVEAIHTDTD------NLGIRIPVGHVDYFVNGGQDQPGCPTFFYAGY-----------
PLA1A-variant3          LFVEAIHTDTD------NLGIRIPVGHVDYFVNGGQDQPGCPTFFYAGY-----------
X2-LIPH                 ------------------LGYKEPLGNIDFYPNGGLDQPGCPKTILGGF-----------
X3-LIPH                 QFVDVIHSDTD------ALGYKEPLGNIDFYPNGGLDQPGCPKTILGGF-----------
LIPH                    QFVDVIHSDTD------ALGYKEPLGNIDFYPNGGLDQPGCPKTILGGF-----------
X1-LIPH                 QFVDVIHSDTD-----------------------------------GF-----------
LIPI-variant7           KFVDVIHSDSN------GLGIQEPLGHIDFYPNGGNKQPGCPKSIFSGI-----------
LIPI-variant.deltaE8    KFVDVIHSDSN------GLGIQEPLGHIDFYPNGGNKQPGCPKSIFSGI-----------
LIPI-variant.deltaE7    KFVDVIHSDSN------GLGIQEPLGHIDFYPNGGNKQPGCPKSIFSGI-----------
LIPI-variant6           KFVDVIHSDSN------GLGIQEPLGHIDFYPNGGNKQPGC--------------------
LIPI-variant2           KFVDVIHSDSN------GLGIQEPLGHIDFYPNGGNKQPGCPKSIFSGI-----------
X1-LIP1                 KFVDVIHSDSN------GLGIQEPLGHIDFYPNGGNKQPGCPKSIFSGI-----------
LIPI-variant.fl         KFVDVIHSDSN------GLGIQEPLGHIDFYPNGGNKQPGCPKSIFSGI-----------
LIPI-variant.deltaE5    KFVDVIHSDSN------G-----------------------------I-----------

X5-LIPC                 TQTIKCSHERSVHLFIDSLLHAGTQSMAYPCGDMNSFSQGLCLSCK---KGRCNTLGYHV
LIPC                    TQTIKCSHERSVHLFIDSLLHAGTQSMAYPCGDMNSFSQGLCLSCK---KGRCNTLGYHV
X1-LIPC                 TQTIKCSHERSVHLFIDSLLHAGTQSMAYPCGDMNSFSQGLCLSCK---KGRCNTLGYHV
X6-LIPC                 TQTIKCSHERSVHLFIDSLLHAGTQSMAYPCGDMNSFSQGLCLSCK---KGRCNTLGYHV
LPL                     DQLVKCSHERSIHLFIDSLLNEENPSKAYRCSSKEAFEKGLCLSCR---KNRCNNLGYEI
LIPG-variant1           TEVVKCEHERAVHLFVDSLVNQDKPSFAFQCTDSNRFKKGICLSCR---KNRCNSIGYNA
X1-LIPG                 TEVVKCEHERAVHLFVDSLVNQDKPSFAFQCTDSNRFKKGICLSCR---KNRCNSIGYNA
X3-LIPG                 TEVVKCEHERAVHLFVDSLVNQDKPSFAFQCTDSNRFKKGICLSCR---KNRCNSIGYNA
X4-LIPG                 TEVVKCEHERAVHLFVDSLVNQDKPSFAFQCTDSNRFKKGICLSCR---KNRCNSIGYNA
X2-LIPG                 TEVVKCEHERAVHLFVDSLVNQDKPSFAFQCTDSNRFKKGICLSCR---KNRCNSIGYNA
LIPG-variant2           TEVVKCEHERAVHLFVDSLVNQDKPSFAFQCTDSNRFKKGICLSCR---KNRCNSIGYNA
X1-PNLIPRP3             ------------------------------------------------------------
PNLIPRP3                ASFFDCNHARSYQFYAESILNPD-AFIAYPCRSYTSFKAGNCFFCS---KEGCPTMGHFA
X3-PNLIPRP3             ASFFDCNHARSYQFYAESILNPD-AFIAYPCRSYTSFKAGNCFFCS---KEGCPTMGHFA
X2-PNLIPRP3             ASFFDCNHARSYQFYAESILNPD-AFIAYPCRSYTSFKAGNCFFCS---KEGCPTMGHFA
PNLIPRP2                GGFVSCNHLRSFEYYSSSVLNPD-GFLGYPCASYDEFQESKCFPCP---AEGCPKMGHYA
PNLIP                   RDFAACNHLRSYKYYTDSIVNPD-GFAGFPCASYNVFTANKCFPCP---SGGCPQMGHYA
PNLIPRP1-variant        RDFVACNHLRSYKYYLESILNPD-GFAAYPCTSYKSFESDKCFPCP---DQGCPQMGHYA
X1-PNLIPRP1             RDFVACNHLRSYKYYLESILNPD-GFAAYPCTSYKSFESDKCFPCP---DQGCPQMGHYA
PLA1A-variant4          -SYLICDHMRAVHLYISALENSC-PLMAFPCASYKAFLAGRCLDCFNPFLLSCPRIGLVE
```

```
PLA1A-variant1          -SYLICDHMRAVHLYISALENSC-PLMAFPCASYKAFLAGRCLDCFNPFLLSCPRIGLVE
PLA1A-variant2          -SYLICDHMRAVHLYISALENSC-PLMAFPCASYKAFLAGRCLDCFNPFLLSCPRIGLVE
PLA1A-variant3          -SYLICDHMRAVHLYISALENSC-PLMAFPCASYKAFLAGRCLDCFNPFLLSCPRIGLVE
X2-LIPH                 -QYFKCDHQRSVYLYLSSLRESC-TITAYPCDSYQDYRNGKCVSCGTSQKESCPLLGYYA
X3-LIPH                 -QYFKCDHQRSVYLYLSSLRESC-TITAYPCDSYQDYRNGKCVSCGTSQKESCPLLGYYA
LIPH                    -QYFKCDHQRSVYLYLSSLRESC-TITAYPCDSYQDYRNGKCVSCGTSQKESCPLLGYYA
X1-LIPH                 -QYFKCDHQRSVYLYLSSLRESC-TITAYPCDSYQDYRNGKCVSCGTSQKESCPLLGYYA
LIPI-variant7           -QFIKCNHQRAVHLFMASLETNC-NFISFPCRSYKDYKTSLCVDCDCFKEKSCPRLGYQA
LIPI-variant.deltaE8    -QFIKCNHQRAVHLFMASLETNC-NFISFPCRSYKDYKTSLCVDCDCFKEKSCPRLGYQA
LIPI-variant.deltaE7    -QFIKCNHQRAVHLFMASLETNC-NFISFPCRSYKDYKTSLCVDCDCFKEKSCPRLGYQA
LIPI-variant6           ------------------------------------------------------------
LIPI-variant2           -QFIKCNHQRAVHLFMASLETNC-NFISFPCRSYKDYKTSLCVDCDCFKEKSCPRLGYQA
X1-LIP1                 -QFIKCNHQRAVHLFMASLETNC-NFISFPCRSYKDYKTSLCVDCDCFKEKSCPRLGYQA
LIPI-variant.fl         -QFIKCNHQRAVHLFMASLETNC-NFISFPCRSYKDYKTSLCVDCDCFKEKSCPRLGYQA
LIPI-variant.deltaE5    -QFIKCNHQRAVHLFMASLETNC-NFISFPCRSYKDYKTSLCVDCDCFKEKSCPRLGYQA


X5-LIPC                 RQEPRSKS---------KRLFLVTRAQSPFKVYHYQFKIQFINQ-TETPIQTTFTMSLLG
LIPC                    RQEPRSKS---------KRLFLVTRAQSPFKVYHYQFKIQFINQ-TETPIQTTFTMSLLG
X1-LIPC                 RQEPRSKS---------KRLFLVTRAQSPFKVYHYQFKIQFINQ-TETPIQTTFTMSLLG
X6-LIPC                 RQEPRSKS---------KRLFLVTRAQSPFKGFQLE------------------------
LPL                     NKVRAKRS---------SKMYLKTRSQMPYKVFHYQVKIHFSGTESETHTNQAFEISLYG
LIPG-variant1           KKMRNKRN---------SKMYLKTRAGMPFRVYHYQMKIHVFSYKNMGEIEPTFYVTLYG
X1-LIPG                 KKMRNKRN---------SKMYLKTRAGMPFRVYHYQMKIHVFSYKNMGEIEPTFYVTLYG
X3-LIPG                 KKMRNKRN---------SKMYLKTRAGMPFRVYHYQMKIHVFSYKNMGEIEPTFYVTLYG
X4-LIPG                 KKMRNKRN---------SKMYLKTRAGMPFRVYHYQMKIHVFSYKNMGEIEPTFYVTLYG
X2-LIPG                 KKMRNKRN---------SKMYLKTRAGMPFRVYHYQMKIHVFSYKNMGEIEPTFYVTLYG
LIPG-variant2           KKMRNKRN---------SKMYLKTRAGMPFRVYHYQMKIHVFSYKNMGEIEPTFYVTLYG
X1-PNLIPRP3             ------------------------------------------------------------
PNLIPRP3                DRFHFKNM-----KTNGSHYFLNTGSLSPFARWRHKLSVKLSGSEVT---QGTVFLRVGG
X3-PNLIPRP3             DRFHFKNM-----KTNGSHYFLNTGSLSPFARWRHKLSVKLSGSEVT---QGTVFLRVGG
X2-PNLIPRP3             DRFHFKNM-----KTNGSHYFLNTGSLSPFARWRHKLSVKLSGSEVT---QGTVFLRVGG
PNLIPRP2                DQFKGKTS------AVEQTFFLNTGESGNFTSWRYKVSVTLSGKEKV---NGYIRIALYG
PNLIP                   DRYPGKTN------DVGQKFYLDTGDASNFARWRYKVSVTLSGKK-V---TGHILVSLFG
PNLIPRP1-variant        DKFAGRTS------EEQQKFFLNTGEASNFARWRYGVSITLSGRT-A---TGQIKVALFG
X1-PNLIPRP1             DKFAGRTS------EEQQKFFLNTGEASNFARWRYGVSITLSGRT-A---TGQIKVALFG
PLA1A-variant4          QG-----GVKIEPLPKEVKVYLLTTSSAPYCMHHSLVEFHL--------------------
PLA1A-variant1          QG-----GVKIEPLPKEVKVYLLTTSSAPYCMHHSLVEFHL--------------------
PLA1A-variant2          QG-----GVKIEPLPKEVKVYLLTTSSAPYCMHHSLVEFHL--------------------
PLA1A-variant3          QG-----GVKIEPLPKEVKVYLLTTSSAPYCMHHSLVEFHL--------------------
X2-LIPH                 DNWKDHLRGK---DPPMTKAFFDTAEESPFCMYHYFVDI----------------------
X3-LIPH                 DNWKDHLRGK---DPPMTKAFFDTAEESPFCMYHYFVDI----------------------
LIPH                    DNWKDHLRGK---DPPMTKAFFDTAEESPFCMYHYFVDI----------------------
X1-LIPH                 DNWKDHLRGK---DPPMTKAFFDTAEESPFCMYHYFVDI----------------------
LIPI-variant7           KLFKGVLKERMEGRPLRTTVFLDTSGTYPFCTYYFVLSI----------------------
LIPI-variant.deltaE8    KLFKGVLKERMEGRPLRTTVFLDTSGTYPFCNHHFAGIILYLKTERK---CFLIQTHVHQ
LIPI-variant.deltaE7    KLFKGVLKERM--------------------------------------------------
LIPI-variant6           ------------------------------------------------------------
LIPI-variant2           KLFKGVLKERMEGRPLRTTVFLDTSGTYPFCTYYFVLSI----------------------
X1-LIP1                 KLFKGVLKERMEGRPLRTTVFLDTSGTYPFCTYYFVLSI----------------------
LIPI-variant.fl         KLFKGVLKERMEGRPLRTTVFLDTSGTYPFCTYYFVLSI----------------------
LIPI-variant.deltaE5    KLFKGVLKERMEGRPLRTTVFLDTSGTYPFCTYYFVLSI----------------------


X5-LIPC                 TKEKMQKIPITLGKGIASNKTYSFLITLDVDIGELIMIKFKWENSA--VWANVWDTVQTI
LIPC                    TKEKMQKIPITLGKGIASNKTYSFLITLDVDIGELIMIKFKWENSA--VWANVWDTVQTI
X1-LIPC                 TKEKMQKIPITLGKGIASNKTYSFLITLDVDIGELIMIKFKWENSA--VWANVWDTVQTI
X6-LIPC                 ------------------------------------------------------------
LPL                     TVAESENIPFTLP-EVSTNKTYSFLIYTEVDIGELLMLKLKWKSDSYFSWSDWWSS----
LIPG-variant1           TNADSQTLPLEIVERIEQNATNTFLVYTEEDLGDLLKIQLTWEGAS-QSWYNLWKEFRSY
X1-LIPG                 TNADSQTLPLEIVERIEQNATNTFLVYTEEDLGDLLKIQLTWEGAS-QSWYNLWKEFRSY
X3-LIPG                 TNADSQTLPLEIVERIEQNATNTFLVYTEEDLGDLLKIQLTWEGAS-QSWYNLWKEFRSY
X4-LIPG                 TNADSQTLPLEIVERIEQNATNTFLVYTEEDLGDLLKIQLTWEGAS-QSWYNLWKEFRSY
X2-LIPG                 TNADSQTLPLEIVERIEQNATNTFLVYTEEDLGDLLKIQLTWEGAS-QSWYNLWKEFRSY
LIPG-variant2           TNADSQTLPLEIVERIEQNATNTFLVYTEEDLGDLLKIQLTWEGAS-QSWYNLWKEFRSY
X1-PNLIPRP3             ------------------------------------------------------------
PNLIPRP3                AVRKTGEFAIVSG-KLEPGMTYTKLIDADVNVGNITSVQFIWKKHL---------------
X3-PNLIPRP3             AVRKTGEFAIVSG-KLEPGMTYTKLIDADVNVGNITSVQFIWKKHL---------------
X2-PNLIPRP3             AVRKTGEFAIVSG-KLEPGMTYTKLIDADVNVGNITSVQFIWKKHL---------------
PNLIPRP2                SNENSKQYEIFKG-SLKPDASHTCAIDVDFNVGKIQKVKFLWNKRG---------------
PNLIP                   NKGNSKQYEIFKG-TLKPDSTHSNEFDSDVDVGDLQMVKFIWYNNV---------------
PNLIPRP1-variant        NKGNTHQYSIFRG-ILKPGSTHSYEFDAKLDVGTIEKVKFLWNNNV---------------
X1-PNLIPRP1             NKGNTHQYSIFRG-ILKPGSTHSYEFDAKLDVGTIEKVKFLWNNNV---------------
PLA1A-variant4          ------------------------------------------------------------
PLA1A-variant1          ------------------------------------------------------------
PLA1A-variant2          ------------------------------------------------------------
PLA1A-variant3          ------------------------------------------------------------
X2-LIPH                 ------------------------------------------------------------
X3-LIPH                 ------------------------------------------------------------
LIPH                    ------------------------------------------------------------
X1-LIPH                 ------------------------------------------------------------
LIPI-variant7           ------------------------------------------------------------
LIPI-variant.deltaE8    ------------------------------------------------------------
LIPI-variant.deltaE7    ------------------------------------------------------------
LIPI-variant6           ------------------------------------------------------------
LIPI-variant2           ------------------------------------------------------------
X1-LIP1                 ------------------------------------------------------------
```

```
LIPI-variant.fl          ------------------------------------------------------------
LIPI-variant.deltaE5     ------------------------------------------------------------


X5-LIPC                  IPWSTGPRHSGLVLKTIRVKAGETQQRMTFCSENTDDLLLRPTQEKIFVKCEIKSKTSK-
LIPC                     IPWSTGPRHSGLVLKTIRVKAGETQQRMTFCSENTDDLLLRPTQEKIFVKCEIKSKTSK-
X1-LIPC                  IPWSTGPRHSGLVLKTIRVKAGETQQRMTFCSENTDDLLLRPTQEKIFVKCEIKSKTSK-
X6-LIPC                  ------------------------------------------------------------
LPL                      ---------PGFAIQKIRVKAGETQKKVIFCSREKVSHLQKGKAPAVFVKCHDKSLNKKS
LIPG-variant1            LSQPRNP-GRELNIRRIRVKSGETQRKLTFCTEDPENTSISPGRELWFRKCRD-------
X1-LIPG                  LSQPRNP-GRELNIRRIRVKSGETQRKLTFCTEDPENTSISPGRELWFRKCRD-------
X3-LIPG                  LSQPRNP-GRELNIRRIRVKSGETQRKLTFCTEDPENTSISPGRELWFRKCRD-------
X4-LIPG                  LSQPRNP-GRELNIRRIRVKSGETQRKLTFCTEDPENTSISPGRELWFRKCRD-------
X2-LIPG                  LSQPRNP-GRELNIRRIRVKSGETQRKLTFCTEDPENTSISPGRELWFRKCRD-------
LIPG-variant2            LSQPRNP-GRELNIRRIRVKSGETQRKLTFCTEDPENTSISPGRELWFRKCRD-------
X1-PNLIPRP3              ------------------------------------------------------------
PNLIPRP3                 ----FEDSQNKLGAEMVINTSGKYGYKSTFCSQD--------------------------
X3-PNLIPRP3              ----FEDSQNKLGAEMVINTSGKYGYKSTFCSQD--------------------------
X2-PNLIPRP3              ----FEDSQNKLGAEMVINTSGKYGYKSTFCSQD--------------------------
PNLIPRP2                 ----INLSEPKLGASQITVQSGEDGTEYNFCSSDTV-------------------------
PNLIP                    ----INPTLPRVGASKIIVETNV-GKQFNFCSPETV-------------------------
PNLIPRP1-variant         ----INPTLPKVGATKITVQKGEEKTVYNFCSEDTV-------------------------
X1-PNLIPRP1              ----INPTLPKVGATKITVQKGEEKTVYNFCSEDTV-------------------------
PLA1A-variant4           ------------------------------------------------------------
PLA1A-variant1           ------------------------------------------------------------
PLA1A-variant2           ------------------------------------------------------------
PLA1A-variant3           ------------------------------------------------------------
X2-LIPH                  ------------------------------------------------------------
X3-LIPH                  ------------------------------------------------------------
LIPH                     ------------------------------------------------------------
X1-LIPH                  ------------------------------------------------------------
LIPI-variant7            ------------------------------------------------------------
LIPI-variant.deltaE8     ------------------------------------------------------------
LIPI-variant.deltaE7     ------------------------------------------------------------
LIPI-variant6            ------------------------------------------------------------
LIPI-variant2            ------------------------------------------------------------
X1-LIP1                  ------------------------------------------------------------
LIPI-variant.fl          ------------------------------------------------------------
LIPI-variant.deltaE5     ------------------------------------------------------------


X5-LIPC                  -
LIPC                     -
X1-LIPC                  -
X6-LIPC                  -
LPL                      G
LIPG-variant1            -
X1-LIPG                      -
X3-LIPG                      -
X4-LIPG                      -
X2-LIPG                      -
LIPG-variant2                -
X1-PNLIPRP3                  -
PNLIPRP3                     -
X3-PNLIPRP3                  -
X2-PNLIPRP3                  -
PNLIPRP2                     -
PNLIP                        -
PNLIPRP1-variant             -
X1-PNLIPRP1                  -
PLA1A-variant4               -
PLA1A-variant1               -
PLA1A-variant2               -
PLA1A-variant3               -
X2-LIPH                      -
X3-LIPH                      -
LIPH                         -
X1-LIPH                      -
LIPI-variant7                -
LIPI-variant.deltaE8         -
LIPI-variant.deltaE7         -
LIPI-variant6                -
LIPI-variant2                -
X1-LIP1                      -
LIPI-variant.fl              -
LIPI-variant.deltaE5         -
```

APPENDIX 3. THE PERCENT IDENTITY MATRIX OF THE ALIGNED SEQUENCES.

	X5-LIPC	LIPC	X1-LIPC	X6-LIPC	LPL	LIPG-variant1	X1-LIPG	X3-LIPG	X4-LIPG	X2-LIPG	LIPG-variant2	X1-PNLIPR	PNLIRP3	X3-PNLIPRP	X2-PNLIPRP	PNLIPRP2	PNLIP	PNLIPRP1-variant1	X1-PNLIPRP1	PLA1A-variant4	PLA1A-variant1	PLA1A-variant2	PLA1A-variant3	X2-LIPH	X3-LIPH	LIPH	X1-LIPH	LIPI-variant7	LIPI-variant.deltaE8	LIPI-variant.deltaE7	LIPI-variant6	LIPI-variant2	X1-LIP1	LIPI-variant.fl	LIPI-variant.deltaE5
X5-LIPC	100	100	100	98,3	47,72	45	45	45,52	42,81	41,91	41,91	46,33	34,68	35,54	35,54	31,25	32,81	32,98	33,22	36,67	36,67	37,8	36,42	36,91	39,7	39,7	37,97	40,13	36,92	40	44,38	39,31	39,31	39,31	37,5
LIPC	100	100	100	98,3	47,72	45	45	45,52	42,81	41,91	41,91	46,33	34,68	35,54	35,54	31,25	32,81	32,98	33,22	36,67	36,67	37,8	36,42	36,91	39,7	39,7	37,97	40,13	36,92	40	44,38	39,31	39,31	39,31	37,5
X1-LIPC	100	100	100	98,3	47,72	45	45	45,52	42,81	41,91	41,91	46,33	34,68	35,54	35,54	31,25	32,81	32,98	33,22	36,67	36,67	37,8	36,42	36,91	39,7	39,7	37,97	40,13	36,92	40	44,38	39,31	39,31	39,31	37,5
X6-LIPC	98,3	98,3	98,3	100	48,45	49,13	49,13	50,19	48,68	45,54	45,54	46,33	38,99	40,93	40,93	34,49	34,62	35,31	36,59	36,98	36,98	38,15	36,94	35,65	38,64	38,64	36,75	39,6	39	40	44,38	39	39	39	37,12
LPL	47,72	47,72	47,72	48,45	100	50,12	50,12	51,41	50,73	45,85	45,85	40,11	30,71	31,71	31,71	29,65	32,15	32,41	32,66	34,03	34,03	35,29	40,74	35,62	38,2	38,2	37,55	37,5	35,71	38,75	43,79	37,02	37,02	37,02	34,91
LIPG-variant1	45	45	45	49,13	50,12	100	100	100	100	99,45	99,45	45,45	32,61	33,54	33,54	31,47	33,5	33,16	31,99	36,97	36,97	38,06	40,12	37,77	40,45	40,45	39,66	40,79	39,64	43,75	49,11	42,37	42,37	42,37	40,09
X1-LIPG	45	45	45	49,13	50,12	100	100	100	100	99,45	99,45	45,45	32,61	33,54	33,54	31,47	33,5	33,16	31,99	36,97	36,97	38,06	40,12	37,77	40,45	40,45	39,66	40,79	39,64	43,75	49,11	42,37	42,37	42,37	40,09
X3-LIPG	45,52	45,52	45,52	50,19	51,41	100	100	100	100	99,39	99,39	47,27	33,05	33,54	33,54	32,23	34,35	33,52	31,99	38,43	38,43	39,75	40,12	38,84	41,47	41,47	40,79	40,79	39,49	43,64	49,09	42,25	42,25	42,25	39,91
X4-LIPG	42,81	42,81	42,81	48,68	50,73	100	100	100	100	100	100	48,15	26,83	26,83	26,83	26,61	28,05	27,53	27,53	37,86	37,86	37,86	37,86	36,21	38,46	38,46	36,28	38,46	34,16	40,5	54	38,46	38,46	38,46	32,74
X2-LIPG	41,91	41,91	41,91	45,54	45,85	99,45	99,45	99,39	100	100	100	40,52	29,59	30,31	30,31	28,12	30,91	31,13	28,7	33,18	33,18	34,31	33,67	36,63	35,96	35,96	35,96	30,68	35,19	39,77	43,36	38,38	38,38	38,38	38,38
LIPG-variant2	41,91	41,91	41,91	45,54	45,85	99,45	99,45	99,39	100	100	100	40,52	29,59	30,31	30,31	28,12	30,91	31,13	28,7	33,18	33,18	34,31	33,67	36,63	35,96	35,96	35,96	30,68	35,19	39,77	43,36	38,38	38,38	38,38	38,38
X1-PNLIPRP3	46,33	46,33	46,33	46,33	40,11	45,45	45,45	47,27	48,15	40,52	40,52	100	100	100	100	56,74	58,43	60,11	66,98	41,52	41,52	43,87	47,89	37,04	42,6	42,6	41,38	57,63	46,95	46,95	46,95	46,95	46,95	46,95	45,39
PNLIRP3	34,68	34,68	34,68	38,99	30,71	32,61	32,61	33,05	26,83	29,59	29,59	100	100	100	100	47,35	47,07	48,54	48,2	35,85	35,85	36,95	35,15	34,05	37,97	37,97	37,29	40,38	37,99	42,13	46,95	40,61	40,61	40,61	39,83
X3-PNLIPRP3	35,54	35,54	35,54	40,93	31,71	33,54	33,54	33,54	26,83	30,31	30,31	100	100	100	100	48,52	48,21	48,96	48,2	36,89	36,89	38,28	35,15	36,08	40,35	40,35	39,9	40,38	38,62	43,56	50,38	41,67	41,67	41,67	40,91
X2-PNLIPRP3	35,54	35,54	35,54	40,93	31,71	33,54	33,54	33,54	26,83	30,31	30,31	100	100	100	100	48,52	48,21	48,96	48,2	36,89	36,89	38,28	35,15	36,08	40,35	40,35	39,9	40,38	38,62	43,56	50,38	41,67	41,67	41,67	40,91
PNLIPRP2	31,25	31,25	31,25	34,49	29,65	31,47	31,47	32,23	26,61	28,12	28,12	56,74	47,35	48,52	48,52	100	64,44	63,46	60,59	35,17	35,17	36,5	36,36	35,32	39,41	39,41	38,91	41,94	37,59	41,42	44,64	39,77	39,77	39,77	38,89
PNLIP	32,81	32,81	32,81	34,62	32,15	33,5	33,5	34,35	28,05	30,91	30,91	58,43	47,07	48,21	48,21	64,44	100	67,25	67,32	35,99	35,99	37	35,15	36,17	40,15	40,15	39,33	40	38,08	42,68	47,02	40,53	40,53	40,53	39,32
PNLIPRP1-variant1	32,98	32,98	32,98	35,31	32,41	33,16	33,16	33,52	27,53	31,13	31,13	60,11	48,54	48,96	48,96	63,46	67,25	100	100	34,03	34,03	34,93	32,93	36,32	39,18	39,18	38,66	38,71	35,36	39,08	41,92	37,64	37,64	37,64	36,48
X1-PNLIPRP1	33,22	33,22	33,22	36,59	32,66	31,99	31,99	31,99	27,53	28,7	28,7	66,98	48,2	48,2	48,2	60,59	67,32	100	100	35,42	35,42	35,14	32,93	37,27	41,03	41,03	40,61	38,71	36,32	41,76	48,48	39,49	39,49	39,49	38,18
PLA1A-variant4	36,67	36,67	36,67	36,98	34,03	36,97	36,97	38,43	37,86	33,18	33,18	41,52	35,85	36,89	36,89	35,17	35,99	34,03	35,42	100	100	100	100	39,24	41,33	41,33	40,66	40,88	43,91	45,64	49,7	43,49	43,49	43,49	42,26
PLA1A-variant1	36,67	36,67	36,67	36,98	34,03	36,97	36,97	38,43	37,86	33,18	33,18	41,52	35,85	36,89	36,89	35,17	35,99	34,03	35,42	100	100	100	100	39,24	41,33	41,33	40,66	40,88	43,91	45,64	49,7	43,49	43,49	43,49	42,26
PLA1A-variant2	37,8	37,8	37,8	38,15	35,29	38,06	38,06	39,75	37,86	34,31	34,31	43,87	36,95	38,28	38,28	36,5	37	34,93	35,14	100	100	100	100	40,72	42,75	42,75	42,22	40,88	45,1	47,11	52,29	44,66	44,66	44,66	43,5
PLA1A-variant3	36,42	36,42	36,42	36,94	40,74	40,12	40,12	40,12	37,86	33,67	33,67	47,89	35,15	35,15	35,15	36,36	35,15	32,93	32,93	100	100	100	100	40,74	43,79	43,79	43,17	40,88	43,68	46,53	56,94	43,02	43,02	43,02	40,85
X2-LIPH	36,91	36,91	36,91	35,65	35,62	37,77	37,77	38,84	36,21	36,63	36,63	37,04	34,05	36,08	36,08	35,32	36,17	36,32	37,27	39,24	39,24	40,72	40,74	100	99,59	99,59	99,53	48,03	53,36	55,19	61,76	53,36	53,36	53,36	50,48
X3-LIPH	39,7	39,7	39,7	38,64	38,2	40,45	40,45	41,47	38,46	35,96	35,96	42,6	37,97	40,35	40,35	39,41	40,15	39,18	41,03	41,33	41,33	42,75	43,79	99,59	100	100	100	52,17	55,15	56,91	62,94	55,15	55,15	55,15	52,89
LIPH	39,7	39,7	39,7	38,64	38,2	40,45	40,45	41,47	38,46	35,96	35,96	42,6	37,97	40,35	40,35	39,41	40,15	39,18	41,03	41,33	41,33	42,75	43,79	99,59	100	100	100	52,17	55,15	56,91	62,94	55,15	55,15	55,15	52,89
X1-LIPH	37,97	37,97	37,97	36,75	37,55	39,66	39,66	40,79	36,28	35,96	35,96	41,38	37,29	39,9	39,9	38,91	39,33	38,66	40,61	40,66	40,66	42,22	43,17	99,53	100	100	100	48,09	53,31	55,09	60,96	53,31	53,31	53,31	53,11
LIPI-variant7	40,13	40,13	40,13	39,6	37,5	40,79	40,79	40,79	38,46	30,68	30,68	57,63	40,38	40,38	40,38	41,94	40	38,71	38,71	40,88	40,88	40,88	40,88	48,03	52,17	52,17	48,09	100	95,73	99,26	98,31	99,39	99,39	99,39	99,25
LIPI-variant.deltaE8	36,92	36,92	36,92	39	35,71	39,64	39,64	39,49	34,16	35,19	35,19	46,95	37,99	38,62	38,62	37,59	38,08	35,36	36,32	43,91	43,91	45,1	43,68	53,36	55,15	55,15	53,31	95,73	100	100	100	97,82	97,82	97,82	97,55
LIPI-variant.deltaE7	40	40	40	40	38,75	43,75	43,75	43,64	40,5	39,77	39,77	46,95	42,13	43,56	43,56	41,42	42,68	39,08	41,76	45,64	45,64	47,11	46,53	55,19	56,91	56,91	55,09	99,26	100	100	100	100	100	100	100
LIPI-variant6	44,38	44,38	44,38	44,38	43,79	49,11	49,11	49,09	54	43,36	43,36	46,95	46,95	50,38	50,38	44,64	47,02	41,92	48,48	49,7	49,7	52,29	56,94	61,76	62,94	62,94	60,96	98,31	100	100	100	100	100	100	100
LIPI-variant2	39,31	39,31	39,31	39	37,02	42,37	42,37	42,25	38,46	38,38	38,38	46,95	40,61	41,67	41,67	39,77	40,53	37,64	39,49	43,49	43,49	44,66	43,02	53,36	55,15	55,15	53,31	99,39	97,82	100	100	100	100	100	100
X1-LIP1	39,31	39,31	39,31	39	37,02	42,37	42,37	42,25	38,46	38,38	38,38	46,95	40,61	41,67	41,67	39,77	40,53	37,64	39,49	43,49	43,49	44,66	43,02	53,36	55,15	55,15	53,31	99,39	97,82	100	100	100	100	100	100
LIPI-variant.fl	39,31	39,31	39,31	39	37,02	42,37	42,37	42,25	38,46	38,38	38,38	46,95	40,61	41,67	41,67	39,77	40,53	37,64	39,49	43,49	43,49	44,66	43,02	53,36	55,15	55,15	53,31	99,39	97,82	100	100	100	100	100	100
LIPI-variant.deltaE5	37,5	37,5	37,5	37,12	34,91	40,09	40,09	39,91	32,74	38,38	38,38	45,39	39,83	40,91	40,91	38,89	39,32	36,48	38,18	42,26	42,26	43,5	40,85	50,48	52,89	52,89	53,11	99,25	97,55	100	100	100	100	100	100

APPENDIX 4. SEQUENCES FOR THE FINAL MSA

```
>LIPC
-------------------------CQIRINHPDTLQECGFNSSLPLVMIIHGWSVDGV
LENWIWQMVAALKSQPAQPVNVGLVDWITLAHDH-YTIAVRNTRLVGKEVAALLRWLEES
VQLSRSHVHLIGYSLGAHVSGFAGSSIGGTHKIGRITGLDAAGPLFEGSAPSNRLSPDDA
NFVDAIHTFTR-EHMGLSVGIKQPIGHYDFYPNGGSFQPGCHFLELYRH--IAQHGFNAI
TQTIKCSHERSVHLFIDSLLHAGTQSMAYPCGDMNSFSQGLCLSCK---KGRCNTLGYHV
RQEPRS---------KSKRLFLVTRAQSPFKVYHYQFKIQFINQ-TETPIQTTFTMSLLG
TKEKMQKIPITLGKGIASNKTYSFLITLDVDIGELIMIKFKWENSA--VWANVWDTVQTI
IPWSTGPRHSGLVLKTIRVKAGETQQRMTFCSENTDDLLLRPTQEKIFVKCEIKSKTSK-
-
>LPL
ADQRRDFIDIESKFALRTPEDTAEDTCHLIPGVAESVATCHFNHSSKTFMVIHGWTVTGM
YESWVPKLVAALYKRE-PDSNVIVVDWLSRAQEH-YPVSAGYTKLVGQDVARFINWMEEE
FNYPLDNVHLLGYSLGAHAAGIAGSLTNK--KVNRITGLDPAGPNFEYAEAPSRLSPDDA
DFVDVLHTFTR-GSPGRSIGIQKPVGHVDIYPNGGTFQPGCNIGEAIRV--IAERGLGDV
DQLVKCSHERSIHLFIDSLLNEENPSKAYRCSSKEAFEKGLCLSCR---KNRCNNLGYEI
NKVRAK---------RSSKMYLKTRSQMPYKVFHYQVKIHFSGTESETHTNQAFEISLYG
TVAESENIPFTLPE-VSTNKTYSFLIYTEVDIGELLMLKLKWKSDSYFSWSDWWSS----
---------PGFAIQKIRVKAGETQKKVIFCSREKVSHLQKGKAPAVFVKCHDKSLNKKS
G
>LIPG-variant1
------------RFNLRTSKDPEHEGCYLSVGHSQPLEDCSFNMTAKTFFIIHGWTMSGI
FENWLHKLVSALHTRE-KDANVVVVDWLPLAHQL-YTDAVNNTRVVGHSIARMLDWLQEK
DDFSLGNVHLIGYSLGAHVAGYAGNFVKG--TVGRITGLDPAGPMFEGADIHKRLSPDDA
DFVDVLHTYTR-S-FGLSIGIQMPVGHIDIYPNGGDFQPGCGLNDVLGS--I---AYGTI
TEVVKCEHERAVHLFVDSLVNQDKPSFAFQCTDSNRFKKGICLSCR---KNRCNSIGYNA
KKMRNK---------RNSKMYLKTRAGMPFRVYHYQMKIHVFSYKNMGEIEPTFYVTLYG
TNADSQTLPLEIVERIEQNATNTFLVYTEEDLGDLLKIQLTWEGASQ-SWYNLWKEFRSY
LSQPRNP-GRELNIRRIRVKSGETQRKLTFCTEDPENTSISPGRELWFRKCRD-------
-
>PNLIPRP3
---------------------------------SSTIQASYFGTDKITRINIAGWKT--D
-GKWQRDMCNVLLQLE--DINCINLDWINGSRE--YIHAVNNLRVVGAEVAYFIDVLMKK
FEYSPSKVHLIGHSLGAHLAGEAGSRIPG---LGRITGLDPAGPFFHNTPKEVRLDPSDA
NFVDVIHTNAARILFELGVGTIDACGHLDFYPNGGKHMPGCEDLITPLLKFNFNAYKKEM
ASFFDCNHARSYQFYAESILNPD-AFIAYPCRSYTSFKAGNCFFCS---KEGCPTMGHFA
DRFHFKNMK-----TNGSHYFLNTGSLSPFARWRHKLSVKLSGSEV---TQGTVFLRVGG
AVRKTGEFAIVSGK-LEPGMTYTKLIDADVNVGNITSVQFIWKKHLF-------------
-----EDSQNKLGAEMVINTSGKYGYKSTFCSQD---------------------------
-
>PNLIPRP2
--------DIDTRFLLYTNENPNNFQ-LITGTEPDTIEASNFQLDRKTRFIIHGFLDKAE
-DSWPSDMCKKMFEVE--KVNCICVDWRHGSRAM-YTQAVQNIRVVGAETAFLIQALSTQ
LGYSLEDVHVIGHSLGAHTAAEAGRRLGG--RVGRITGLDPAGPCFQDEPEEVRLDPSDA
VFVDVIHTDSSPIVPSLGFGMSQKVGHLDFFPNGGKEMPGCKKNVLSTI-TDIDGIWEGI
GGFVSCNHLRSFEYYSSSVLNPD-GFLGYPCASYDEFQESKCFPCP---AEGCPKMGHYA
DQFKGKTS------AVEQTFFLNTGESGNFTSWRYKVSVTLSGKEK---VNGYIRIALYG
SNENSKQYEIFKGS-LKPDASHTCAIDVDFNVGKIQKVKFLWNKRGI-------------
-----NLSEPKLGASQITVQSGEDGTEYNFCSSDTV-------------------------
-
>PNLIP
--------DVNTRFLLYTNENPNNFQ-EVA-ADSSSISGSNFKTNRKTRFIIHGFIDKGE
-ENWLANVCKNLFKVE--SVNCICVDWKGGSRTG-YTQASQNIRIVGAEVAYFVEFLQSA
FGYSPSNVHVIGHSLGAHAAGEAGRRTNG--TIGRITGLDPAEPCFQGTPELVRLDPSDA
KFVDVIHTDGAPIVPNLGFGMSQVVGHLDFFPNGGVEMPGCKKNILSQI-VDIDGIWEGT
RDFAACNHLRSYKYYTDSIVNPD-GFAGFPCASYNVFTANKCFPCP---SGGCPQMGHYA
DRYPGKTN------DVGQKFYLDTGDASNFARWRYKVSVTLSGKK----VTGHILVSLFG
NKGNSKQYEIFKGT-LKPDSTHSNEFDSDVDVGDLQMVKFIWYNNVI-------------
-----NPTLPRVGASKIIVETNV-GKQFNFCSPETV-------------------------
-
>PNLIPRP1-variant1
---------IGTRFLLYTNENPNNFQ-ILLLSDPSTIEASNFQMDRKTRFIIHGFIDKGD
-ESWVTDMCKKLFEVE--EVNCICVDWKKGSQAT-YTQAANNVRVVGAQVAQMLDILLTE
YSYPPSKVHLIGHSLGAHVAGEAGSKTPG---LSRITGLDPVEASFESTPEEVRLDPSDA
DFVDVIHTDAAPLIPFLGFGTNQQMGHLDFFPNGGESMPGCKKNALSQI-VDLDGIWAGT
RDFVACNHLRSYKYYLESILNPD-GFAAYPCTSYKSFESDKCFPCP---DQGCPQMGHYA
DKFAGRTS------EEQQKFFLNTGEASNFARWRYGVSITLSGRT----ATGQIKVALFG
NKGNTHQYSIFRGI-LKPGSTHSYEFDAKLDVGTIEKVKFLWNNNVI-------------
-----NPTLPKVGATKITVQKGEEKTVYNFCSEDTV-------------------------
-
```

```
>PLA1A-variant1
--------DLKVQFLLFVPSNPSCGQ-LVEG--SSDLQNSGFNATLGTKLIIHGFRVLGT
KPSWIDTFIRTLLRAT--NANVIAVDWIYGSTGV-YFSAVKNVIKLSLEISLFLNKLL-V
LGVSESSIHIIGVSLGAHVGGMVGQLFGG--QLGQITGLDPAGPEYTRASVEERLDAGDA
LFVEAIHTDTD------NLGIRIPVGHVDYFVNGGQDQPGCPTFFYAGY-----------
-SYLICDHMRAVHLYISALENSC-PLMAFPCASYKAFLAGRCLDCFNPFLLSCPRIGLVE
QG-----GVKIEPLPKEVKVYLLTTSSAPYCMHHSLVEFHL-------------------
------------------------------------------------------------
------------------------------------------------------------
-
>LIPH
-----------------------------------SSAFGNLNVTKKTTFIVHGFRPTGS
PPVWMDDLVKGLLSVE--DMNVVVVDWNRGATTLIYTHASSKTRKVAMVLKEFIDQML-A
EGASLDDIYMIGVSLGAHISGFVGEMYDG--WLGRITGLDPAGPLFNGKPHQDRLDPSDA
QFVDVIHSDTD------ALGYKEPLGNIDFYPNGGLDQPGCPKTILGGF-----------
-QYFKCDHQRSVYLYLSSLRESC-TITAYPCDSYQDYRNGKCVSCGTSQKESCPLLGYYA
DNWKDHLRGK---DPPMTKAFFDTAEESPFCMYHYFVDI---------------------
------------------------------------------------------------
------------------------------------------------------------
-
>LIPI-variant.fl
----------------------------------------NFNTQKKTVWLIHGYRPVGS
IPLWLQNFVRILLNEE--DMNVIVVDWSRGATTFIYNRAVKNTRKVAVSLSVHIKNLL-K
HGASLDNFHFIGVSLGAHISGFVGKIFHG--QLGRITGLDPAGPRFSRKPPYSRLDYTDA
KFVDVIHSDSN------GLGIQEPLGHIDFYPNGGNKQPGCPKSIFSGI-----------
-QFIKCNHQRAVHLFMASLETNC-NFISFPCRSYKDYKTSLCVDCDCFKEKSCPRLGYQA
KLFKGVLKERMEGRPLRTTVFLDTSGTYPFCTYYFVLSI---------------------
------------------------------------------------------------
------------------------------------------------------------
-
```

APPENDIX 5. THE FINAL MSA.

```
CLUSTAL O(1.2.4) multiple sequence alignment


LIPC                ---------------------------CQIRINHPDTLQECGFNSSLPLVMIIHGWSVDGV
LPL                 ADQRRDFIDIESKFALRTPEDTAEDTCHLIPGVAESVATCHFNHSSKTFMVIHGWTVTGM
LIPG-variant1       ------------RFNLRTSKDPEHEGCYLSVGHSQPLEDCSFNMTAKTFFIIHGWTMSGI
PNLIPRP3            ---------------------------------SSTIQASYFGTDKITRINIAGWKTDG-
PNLIPRP2            --------DIDTRFLLYTNENPNNFQLIT-GTEPDTIEASNFQLDRKTRFIIHGFLDKAE
PNLIP               --------DVNTRFLLYTNENPNNFQEVA--ADSSSISGSNFKTNRKTRFIIHGFIDKGE
PNLIPRP1-variant    ---------IGTRFLLYTNENPNNFQILL-LSDPSTIEASNFQMDRKTRFIIHGFIDKGD
PLA1A-variant1      --------DLKVQFLLFVPSNP---SCGQLVEGSSDLQNSGFNATLGTKLIIHGFRVLGT
LIPH                ------------------------------------SSAFGNLNVTKKTTFIVHGFRPTGS
LIPI-variant.fl     ----------------------------------------NFNTQKKTVWLIHGYRPVGS
                                                           :         : *:   .

LIPC                LENWIWQMVAALKSQPAQPVNVGLVDWITLAHD-HYTIAVRNTRLVGKEVAALLRWLEES
LPL                 YESWVPKLVAALYKRE-PDSNVIVVDWLSRAQE-HYPVSAGYTKLVGQDVARFINWMEEE
LIPG-variant1       FENWLHKLVSALHTRE-KDANVVVVDWLPLAHQ-LYTDAVNNTRVVGHSIARMLDWLQEK
PNLIPRP3            --KWQRDMCNVLLQLE--DINCINLDWINGSR--EYIHAVNNLRVVGAEVAYFIDVLMKK
PNLIPRP2            -DSWPSDMCKKMFEVE--KVNCICVDWRHGSRA-MYTQAVQNIRVVGAETAFLIQALSTQ
PNLIP               -ENWLANVCKNLFKVE--SVNCICVDWKGGSRT-GYTQASQNIRIVGAEVAYFVEFLQSA
PNLIPRP1-variant    -ESWVTDMCKKLFEVE--EVNCICVDWKKGSQA-TYTQAANNVRVVGAQVAQMLDILLTE
PLA1A-variant1      KPSWIDTFIRTLLRAT--NANVIAVDWIYGSTG-VYFSAVKNVIKLSLEISLFLNKLLV-
LIPH                PPVWMDDLVKGLLSVE--DMNVVVVDWNRGATTLIYTHASSKTRKVAMVLKEFIDQMLA-
LIPI-variant.fl     IPLWLQNFVRILLNEE--DMNVIVVDWSRGATTFIYNRAVKNTRKVAVSLSVHIKNLLK-
                       *   .   :        *   :**   :    *  :      :.      :  :

LIPC                VQLSRSHVHLIGYSLGAHVSGFAGSSIGGTHKIGRITGLDAAGPLFEGSAPSNRLSPDDA
LPL                 FNYPLDNVHLLGYSLGAHAAGIAGSLTNK--KVNRITGLDPAGPNFEYAEAPSRLSPDDA
LIPG-variant1       DDFSLGNVHLIGYSLGAHVAGYAGNFVKG--TVGRITGLDPAGPMFEGADIHKRLSPDDA
PNLIPRP3            FEYSPSKVHLIGHSLGAHLAGEAGSRIPG---LGRITGLDPAGPFFHNTPKEVRLDPSDA
PNLIPRP2            LGYSLEDVHVIGHSLGAHTAAEAGRRLGG--RVGRITGLDPAGPCFQDEPEEVRLDPSDA
PNLIP               FGYSPSNVHVIGHSLGAHAAGEAGRRTNG--TIGRITGLDPAEPCFQGTPELVRLDPSDA
PNLIPRP1-variant    YSYPPSKVHLIGHSLGAHVAGEAGSKTPG---LSRITGLDPVEASFESTPEEVRLDPSDA
PLA1A-variant1      LGVSESSIHIIGVSLGAHVGGMVGQLFGG--QLGQITGLDPAGPEYTRASVEERLDAGDA
LIPH                EGASLDDIYMIGVSLGAHISGFVGEMYDG--WLGRITGLDPAGPLFNGKPHQDRLDPSDA
LIPI-variant.fl     HGASLDNFHFIGVSLGAHISGFVGKIFHG--QLGRITGLDPAGPRFSRKPPYSRLDYTDA
                           .:.:* ***** .. .*        :.:***** .   :       **.  **

LIPC                NFVDAIHTFTR-EHMGLSVGIKQPIGHYDFYPNGGSFQPGCHFLELYRH--IAQHGFNAI
LPL                 DFVDVLHTFTR-GSPGRSIGIQKPVGHVDIYPNGGTFQPGCNIGEAIRV--IAERGLGDV
LIPG-variant1       DFVDVLHTYTR--SFGLSIGIQMPVGHIDIYPNGGDFQPGCGLNDVLGS-----IAYGTI
PNLIPRP3            NFVDVIHTNAARILFELGVGTIDACGHLDFYPNGGKHMPGCEDLITPLLKFNFNAYKKEM
PNLIPRP2            VFVDVIHTDSSPIVPSLGFGMSQKVGHLDFFPNGGKEMPGCKKNVLSTI-TDIDGIWEGI
PNLIP               KFVDVIHTDGAPIVPNLGFGMSQVVGHLDFFPNGGVEMPGCKKNILSQI-VDIDGIWEGT
PNLIPRP1-variant    DFVDVIHTDAAPLIPFLGFGTNQQMGHLDFFPNGGESMPGCKKNALSQI-VDLDGIWAGT
PLA1A-variant1      LFVEAIHTDTD------NLGIRIPVGHVDYFVNGGQDQPGCPTFFYAGY-----------
LIPH                QFVDVIHSDTD------ALGYKEPLGNIDFYPNGGLDQPGCPKTILGGF-----------
LIPI-variant.fl     KFVDVIHSDSN------GLGIQEPLGHIDFYPNGGNKQPGCPKSIFSGI-----------
                     **:.:*:          .*     *: * : ***   ***

LIPC                TQTIKCSHERSVHLFIDSLLHAGTQSMAYPCGDMNSFSQGLCLSCK---KGRCNTLGYHV
LPL                 DQLVKCSHERSIHLFIDSLLNEENPSKAYRCSSKEAFEKGLCLSCR---KNRCNNLGYEI
LIPG-variant1       TEVVKCEHERAVHLFVDSLVNQDKPSFAFQCTDSNRFKKGICLSCR---KNRCNSIGYNA
PNLIPRP3            ASFFDCNHARSYQFYAESILNPD-AFIAYPCRSYTSFKAGNCFFCS---KEGCPTMGHFA
PNLIPRP2            GGFVSCNHLRSFEYYSSSVLNPD-GFLGYPCASYDEFQESKCFPCP---AEGCPKMGHYA
PNLIP               RDFAACNHLRSYKYYTDSIVNPD-GFAGFPCASYNVFTANKCFPCP---SGGCPQMGHYA
PNLIPRP1-variant    RDFVACNHLRSYKYYLESILNPD-GFAAYPCTSYKSFESDKCFPCP---DQGCPQMGHYA
PLA1A-variant1      -SYLICDHMRAVHLYISALENSC-PLMAFPCASYKAFLAGRCLDCFNPFLLSCPRIGLVE
LIPH                -QYFKCDHQRSVYLYLSSLRESC-TITAYPCDSYQDYRNGKCVSCGTSQKESCPLLGYYA
LIPI-variant.fl     -QFIKCNHQRAVHLFMASLETNC-NFISFPCRSYKDYKTSLCVDCDCFKEKSCPRLGYQA
                         *.* *:   :  ::        .: * .   :  . *. *       *  :*

LIPC                RQEP---------RSKSKRLFLVTRAQSPFKVYHYQFKIQFINQ-TETPIQTTFTMSLLG
LPL                 NKVR---------AKRSSKMYLKTRSQMPYKVFHYQVKIHFSGTESETHTNQAFEISLYG
LIPG-variant1       KKMR---------NKRNSKMYLKTRAGMPFRVYHYQMKIHVFSYKNMGEIEPTFYVTLYG
PNLIPRP3            DRFHFKNM-----KTNGSHYFLNTGSLSPFARWRHKLSVKLSGSE---VTQGTVFLRVGG
PNLIPRP2            DQFKGKTS------AVEQTFFLNTGESGNFTSWRYKVSVTLSGKE---KVNGYIRIALYG
PNLIP               DRYPGKTN------DVGQKFYLDTGDASNFARWRYKVSVTLSGKK----VTGHILVSLFG
PNLIPRP1-variant    DKFAGRTS------EEQQKFFLNTGEASNFARWRYGVSITLSGRT----ATGQIKVALFG
PLA1A-variant1      QG-----GVKIEPLPKEVKVYLLTTSSAPYCMHHSLVEFHL--------------------
LIPH                DNWKDHLRGK---DPPMTKAFFDTAEESPFCMYHYFVDI----------------------
LIPI-variant.fl     KLFKGVLKERMEGRPLRTTVFLDTSGTYPFCTYYFVLSI----------------------
                                    :: *     :      ...

LIPC                TKEKMQKIPITLGKGIASNKTYSFLITLDVDIGELIMIKFKWENSA--VWANVWDTVQTI
```

```
LPL                 TVAESENIPFTLPE-VSTNKTYSFLIYTEVDIGELLMLKLKWKSDSYFSWSDWWSS----
LIPG-variant1       TNADSQTLPLEIVERIEQNATNTFLVYTEEDLGDLLKIQLTWEGAS-QSWYNLWKEFRSY
PNLIPRP3            AVRKTGEFAIVSGK-LEPGMTYTKLIDADVNVGNITSVQFIWKKHL--------------
PNLIPRP2            SNENSKQYEIFKGS-LKPDASHTCAIDVDFNVGKIQKVKFLWNKRG--------------
PNLIP               NKGNSKQYEIFKGT-LKPDSTHSNEFDSDVDVGDLQMVKFIWYNNV--------------
PNLIPRP1-variant    NKGNTHQYSIFRGI-LKPGSTHSYEFDAKLDVGTIEKVKFLWNNNV--------------
PLA1A-variant1      ------------------------------------------------------------
LIPH                ------------------------------------------------------------
LIPI-variant.fl     ------------------------------------------------------------

LIPC                IPWSTGPRHSGLVLKTIRVKAGETQQRMTFCSENTDDLLLRPTQEKIFVKCEIKSKTSK-
LPL                 ---------PGFAIQKIRVKAGETQKKVIFCSREKVSHLQKGKAPAVFVKCHDKSLNKKS
LIPG-variant1       LSQPRNP-GRELNIRRIRVKSGETQRKLTFCTEDPENTSISPGRELWFRKCRD--------
PNLIPRP3            ----FEDSQNKLGAEMVINTSGKYGYKSTFCSQD-------------------------
PNLIPRP2            ----INLSEPKLGASQITVQSGEDGTEYNFCSSDTV-----------------------
PNLIP               ----INPTLPRVGASKIIVETNV-GKQFNFCSPETV-----------------------
PNLIPRP1-variant    ----INPTLPKVGATKITVQKGEEKTVYNFCSEDTV-----------------------
PLA1A-variant1      ------------------------------------------------------------
LIPH                ------------------------------------------------------------
LIPI-variant.fl     ------------------------------------------------------------

LIPC                -
LPL                 G
LIPG-variant1       -
PNLIPRP3            -
PNLIPRP2            -
PNLIP               -
PNLIPRP1-variant    -
PLA1A-variant1      -
LIPH                -
LIPI-variant.fl     -
```