Genes in general

Computational Biology camp for high school students

January 29, 2012

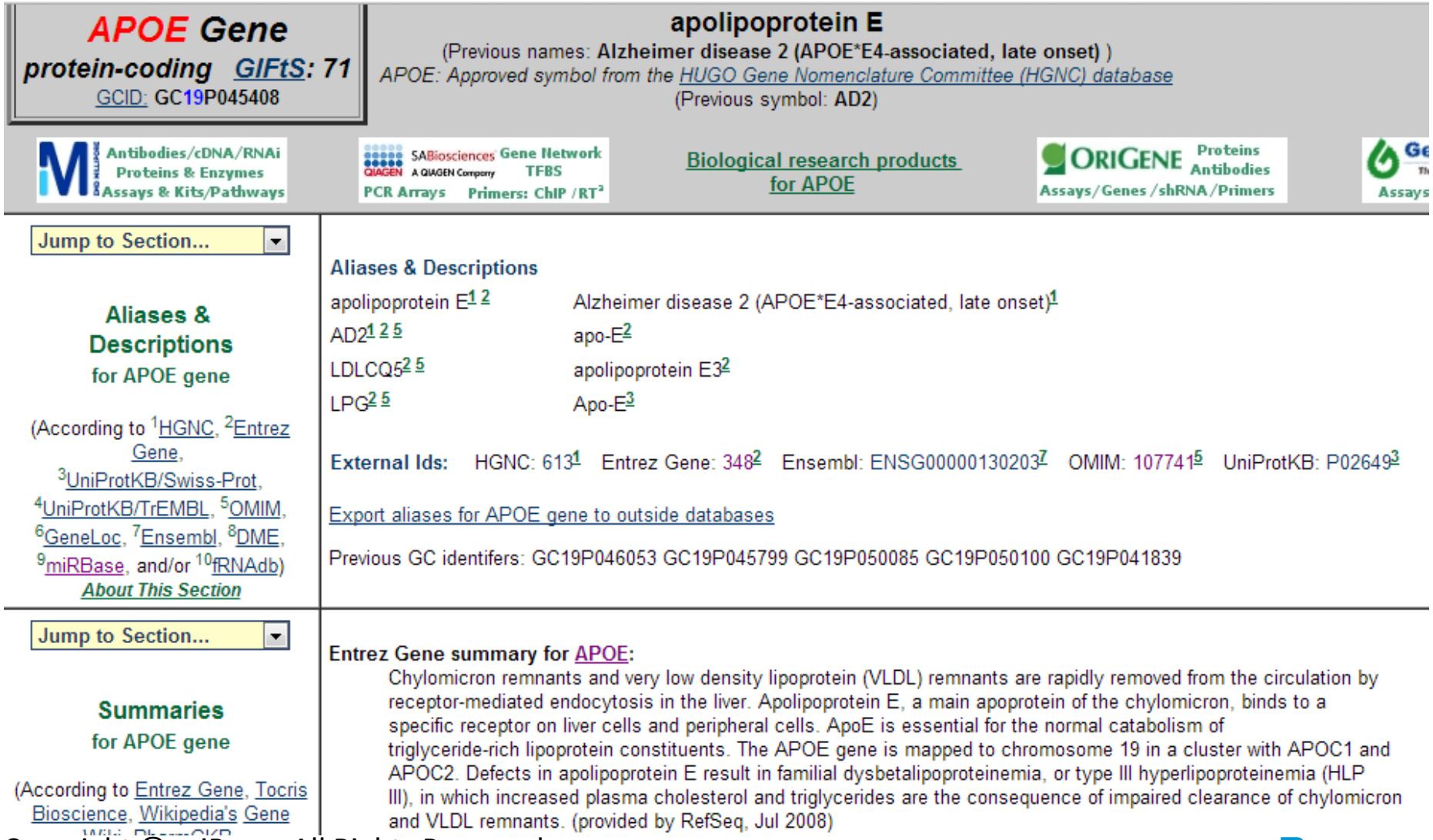
MIRcore

Inhan Lee, Ph.D.
inhan@mircore.org
http://mircore.org

GeneCards
The Human Gene Conpendium
http://www.genecards.org/

MIRCORE

http://www.genecards.org/cgi-bin/carddisp.pl?gene=APOE



APOE Gene
protein-coding GIFtS: 71
GCID: GC19P045408

apolipoprotein E
(Previous names: **Alzheimer disease 2 (APOE*E4-associated, late onset)**)
APOE: Approved symbol from the HUGO Gene Nomenclature Committee (HGNC) database
(Previous symbol: **AD2**)

Jump to Section...

Aliases & Descriptions
for APOE gene

(According to [1]HGNC, [2]Entrez Gene, [3]UniProtKB/Swiss-Prot, [4]UniProtKB/TrEMBL, [5]OMIM, [6]GeneLoc, [7]Ensembl, [8]DME, [9]miRBase, and/or [10]fRNAdb)
About This Section

Aliases & Descriptions

apolipoprotein E[1 2]	Alzheimer disease 2 (APOE*E4-associated, late onset)[1]
AD2[1 2 5]	apo-E[2]
LDLCQ5[2 5]	apolipoprotein E3[2]
LPG[2 5]	Apo-E[3]

External Ids: HGNC: 613[1] Entrez Gene: 348[2] Ensembl: ENSG00000130203[7] OMIM: 107741[5] UniProtKB: P02649[3]

Export aliases for APOE gene to outside databases

Previous GC identifers: GC19P046053 GC19P045799 GC19P050085 GC19P050100 GC19P041839

Jump to Section...

Summaries
for APOE gene

(According to Entrez Gene, Tocris Bioscience, Wikipedia's Gene

Entrez Gene summary for APOE:
Chylomicron remnants and very low density lipoprotein (VLDL) remnants are rapidly removed from the circulation by receptor-mediated endocytosis in the liver. Apolipoprotein E, a main apoprotein of the chylomicron, binds to a specific receptor on liver cells and peripheral cells. ApoE is essential for the normal catabolism of triglyceride-rich lipoprotein constituents. The APOE gene is mapped to chromosome 19 in a cluster with APOC1 and APOC2. Defects in apolipoprotein E result in familial dysbetalipoproteinemia, or type III hyperlipoproteinemia (HLP III), in which increased plasma cholesterol and triglycerides are the consequence of impaired clearance of chylomicron and VLDL remnants. (provided by RefSeq, Jul 2008)

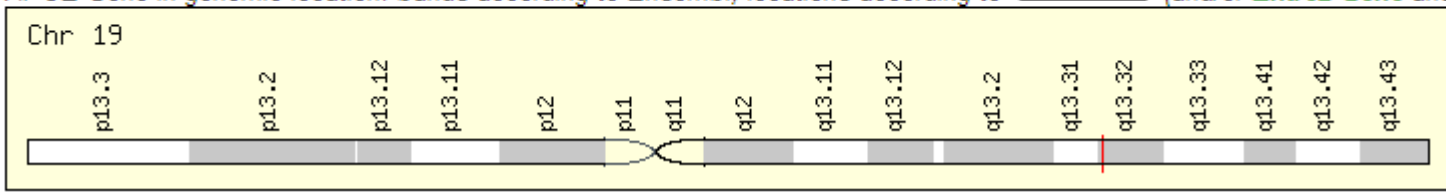
MIRCORE

APOE

Genomic Location:
Genomic View: UCSC Golden Path with GeneCards custom track

Entrez Gene cytogenetic band: **19q13.2** *Ensembl cytogenetic band:* **19q13.32** *HGNC cytogenetic band:* **19q13.31**

APOE Gene in genomic location: bands according to Ensembl, locations according to GeneLoc *(and/or* ***Entrez Gene*** *and*



GeneLoc gene densities for chromosome 19 GeneLoc Exon Structure

GeneLoc location for GC19P045408: *view genomic region* *(about GC identifiers)*

Start: **45,408,956** bp from *pter*
End: **45,412,650** bp from *pter*
Size: **3,695** bases
Orientation: **plus** strand

RefSeq DNA sequence:
NC_000019.9 NT_011109.16

MIRCORE

USCS genome browser

http://genome.ucsc.edu/cgi-bin/hgGateway

Genomes | Genome Browser | Tools | Mirrors | Downloads | My Data | About Us | Help

Human (*Homo sapiens*) Genome Browser Gateway

The UCSC Genome Browser was created by the Genome Bioinformatics Group of UC Santa Cruz.
Software Copyright (c) The Regents of the University of California. All rights reserved.

group	genome	assembly	position	search term
Mammal	Human	Feb. 2009 (GRCh37/hg19)	chr16:15,065,716-15,142,715	enter position, gene symbol or search terms — submit

Click here to reset the browser user interface settings to their defaults.

track search | add custom tracks | track hubs | configure tracks and display

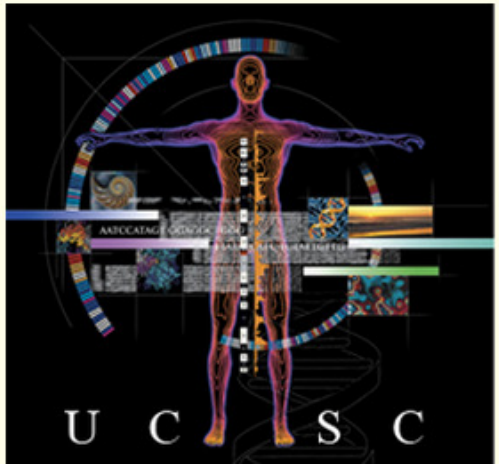
Human Genome Browser – hg19 assembly (sequences)

The February 2009 human reference sequence (GRCh37) was produced by the Genome Reference Consortium. For more information about this assembly, see GRCh37 in the NCBI Assembly database.

Sample position queries

A genome position can be specified by the accession number of a sequenced genomic clone, an mRNA or EST or STS marker, a chromosomal coordinate range, or keywords from the GenBank description of an mRNA. The following list shows examples of valid position queries for the human genome. See the User's Guide for more information.

Request:	Genome Browser Response:
chr7	Displays all of chromosome 7



MIRCORE

Choose RefSeq APOE

UCSC Genes

APOE (uc002pab.3) at chr19:45409039-45412650 - Homo sapiens apolipoprotein E (APOE), mRNA.
TMCC2 (uc021pia.1) at chr1:205197038-205242471 - Homo sapiens transmembrane and coiled-coil domain family 2 (TMCC2),
TMCC2 (uc010prf.2) at chr1:205198059-205242471 - Homo sapiens transmembrane and coiled-coil domain family 2 (TMCC2),
TMCC2 (uc001hcb.2) at chr1:205225329-205242471 - Homo sapiens transmembrane and coiled-coil domain family 2 (TMCC2),
TMCC2 (uc001hca.3) at chr1:205215725-205242471 - Homo sapiens transmembrane and coiled-coil domain family 2 (TMCC2),
LRP10 (uc001whd.3) at chr14:23340960-23347291 - Homo sapiens low density lipoprotein receptor-related protein 10 (LR
LRP10 (uc001whe.3) at chr14:23344264-23349586 - Homo sapiens low density lipoprotein receptor-related protein 10 (LR
ECSIT (uc010dyc.2) at chr19:11616731-11639680 - Homo sapiens ECSIT homolog (Drosophila) (ECSIT), nuclear gene encodi
ECSIT (uc002msb.3) at chr19:11616731-11639987 - Homo sapiens ECSIT homolog (Drosophila) (ECSIT), nuclear gene encodi
APOBR (uc010byq.2) at chr16:28505970-28510291 - Homo sapiens apolipoprotein B receptor (APOBR), mRNA.
APOC1 (uc002pae.1) at chr19:45417921-45422606 - Homo sapiens apolipoprotein C-I (APOC1), mRNA.
APOC1 (uc002pad.1) at chr19:45417577-45422606 - Homo sapiens apolipoprotein C-I (APOC1), mRNA.
APOC1 (uc002pac.1) at chr19:45417577-45422606 - Homo sapiens apolipoprotein C-I (APOC1), mRNA.
LCAT (uc002euy.1) at chr16:67973787-67978015 - Homo sapiens lecithin-cholesterol acyltransferase (LCAT), mRNA.
CDC37 (uc002mof.1) at chr19:10501809-10514271 - Homo sapiens cell division cycle 37 homolog (S. cerevisiae) (CDC37),
ST13 (uc003aze.3) at chr22:41220601-41252687 - Homo sapiens suppression of tumorigenicity 13 (colon carcinoma) (Hsp7
APOBR (uc002dqb.2) at chr16:28505970-28510291 - Homo sapiens apolipoprotein B receptor (APOBR), mRNA.
PDCD4 (uc001kzh.3) at chr10:112631553-112659764 - Homo sapiens programmed cell death 4 (neoplastic transformation in
LRP8 (uc001cvl.2) at chr1:53708041-53793821 - Homo sapiens low density lipoprotein receptor-related protein 8, apoli
LRP8 (uc001cvk.2) at chr1:53708041-53793821 - Homo sapiens low density lipoprotein receptor-related protein 8, apoli
LRP8 (uc001cvj.2) at chr1:53708041-53793821 - Homo sapiens low density lipoprotein receptor-related protein 8, apoli
LRP8 (uc001cvi.2) at chr1:53708041-53793821 - Homo sapiens low density lipoprotein receptor-related protein 8, apoli
LRP8 (uc001cvm.1) at chr1:53716362-53734270 - Homo sapiens low density lipoprotein receptor-related protein 8, apoli
RELN (uc022ajq.1) at chr7:103112231-103629963 - Homo sapiens reelin (RELN), transcript variant 1, mRNA.
RELN (uc022ajr.1) at chr7:103112231-103629963 - Homo sapiens reelin (RELN), transcript variant 1, mRNA.
MAPK8IP2 (uc003bmx.3) at chr22:51039131-51049979 - Homo sapiens mitogen-activated protein kinase 8 interacting prote
MAPK8IP1 (uc001nbr.3) at chr11:45907047-45928016 - Homo sapiens mitogen-activated protein kinase 8 interacting prote

RefSeq Genes

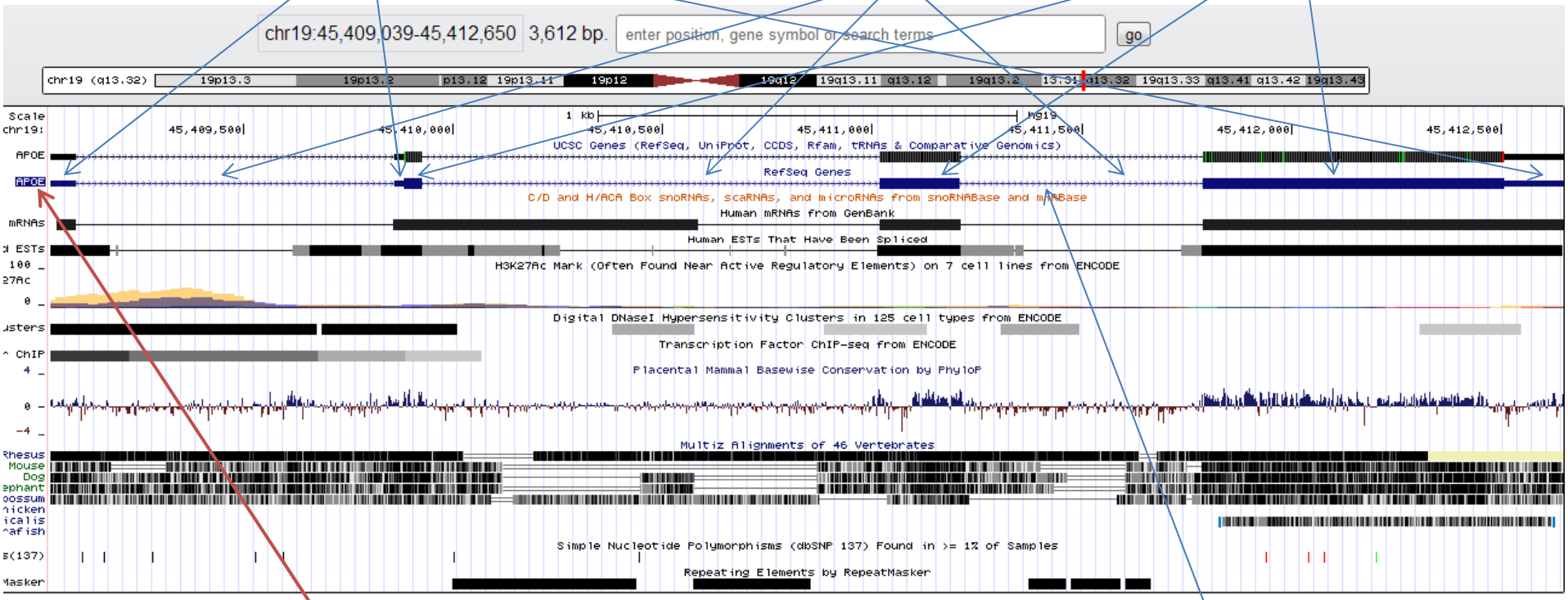
APOE at chr19:45409039-45412650 - (NM_000041) apolipoprotein E precursor

APOE

Thin blocks: UTR (untranslated regions)

Lines: intron

Thick blocks: codon region



Click any of the APOE RefSeq area

Arrow >>>>> shows: positive strand

miRcore

Detailed information

RefSeq Gene APOE

RefSeq: NM_000041.2 **Status:** Reviewed
Description: Homo sapiens apolipoprotein E (APOE), mRNA.
CCDS: CCDS12647.1
CDS: 3' complete
OMIM: 107741
Entrez Gene: 348
PubMed on Gene: APOE
PubMed on Product: apolipoprotein E precursor
GeneCards: APOE
AceView: APOE
Stanford SOURCE: NM_000041

Summary of APOE

Chylomicron remnants and very low density lipoprotein (VLDL) remnants are rapidly removed from the circulation by receptor-mediated endocytosis in the liver. Apolipoprotein E, a main apoprotein of the chylomicron, binds to a specific receptor on liver cells and peripheral cells. ApoE is essential for the normal catabolism of triglyceride-rich lipoprotein constituents. The APOE gene is mapped to chromosome 19 in a cluster with APOC1 and APOC2. Defects in apolipoprotein E result in familial dysbetalipoproteinemia, or type III hyperlipoproteinemia (HLP III), in which increased plasma cholesterol and triglycerides are the consequence of impaired clearance of chylomicron and VLDL remnants. [provided by RefSeq, Jul 2008]. Publication Note: This RefSeq record includes a subset of the publications that are available for this gene. Please see the Gene record to access additional publications. ##RefSeq-Attributes-START## Transcript_exon_combination_evidence :: BC003557.1, M12529.1 [ECO:0000332] ##RefSeq-Attributes-END##

---

mRNA/Genomic Alignments

BROWSER	SIZE	IDENTITY	CHROMOSOME	STRAND	START	END	QUERY	START	END	TOTAL
browser	1180	100.0%	19	+	45409039	45412650	NM_000041	1	1180	1223

MIRCORE

Example tasks

b) Identify the function of the gene (helpful sites: http://www.genecards.org, http://www.ncbi.nlm.nih.gov/gene)

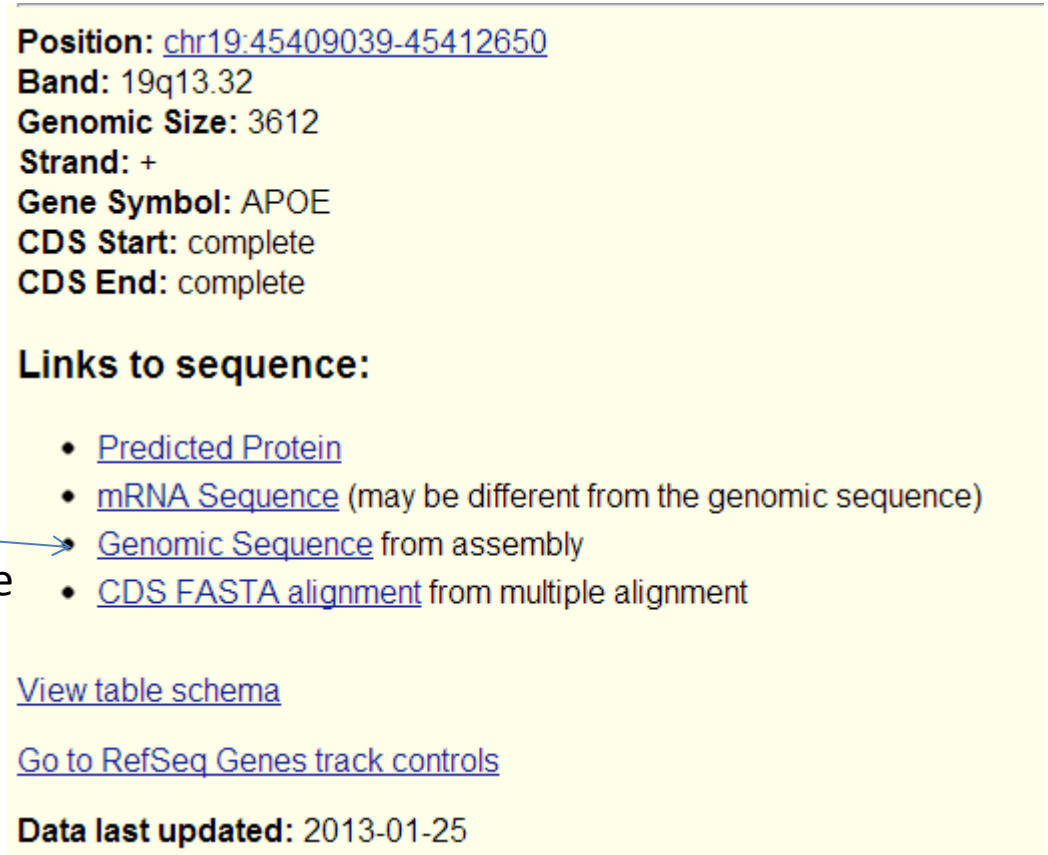
c) Find its genome position (helpful sites: http://genome.ucsc.edu/cgi-bin/hgGateway, http://www.ncbi.nlm.nih.gov/gene)

d) What is the total number of transcribed sequences?

e) Find the mRNA sequences. No need to worry about U/T conversion or upper/lower cases. T will be considered as U (helpful sites:http://genome.ucsc.edu/cgi-bin/hgGateway, http://www.ncbi.nlm.nih.gov/nuccore).

f) Find only coding sequences among the mRNA sequences. No need to worry about U/T conversion or upper/lower cases. T will be considered as U (helpful sites: http://genome.ucsc.edu/cgi-bin/hgGateway: use "Genomic Sequence" rather than "mRNA" sequence for this task, http://www.ncbi.nlm.nih.gov/nuccore).

Let’s get all exon sequences



Position: chr19:45409039-45412650
Band: 19q13.32
Genomic Size: 3612
Strand: +
Gene Symbol: APOE
CDS Start: complete
CDS End: complete

Links to sequence:

- Predicted Protein
- mRNA Sequence (may be different from the genomic sequence)
- Genomic Sequence from assembly
- CDS FASTA alignment from multiple alignment

View table schema

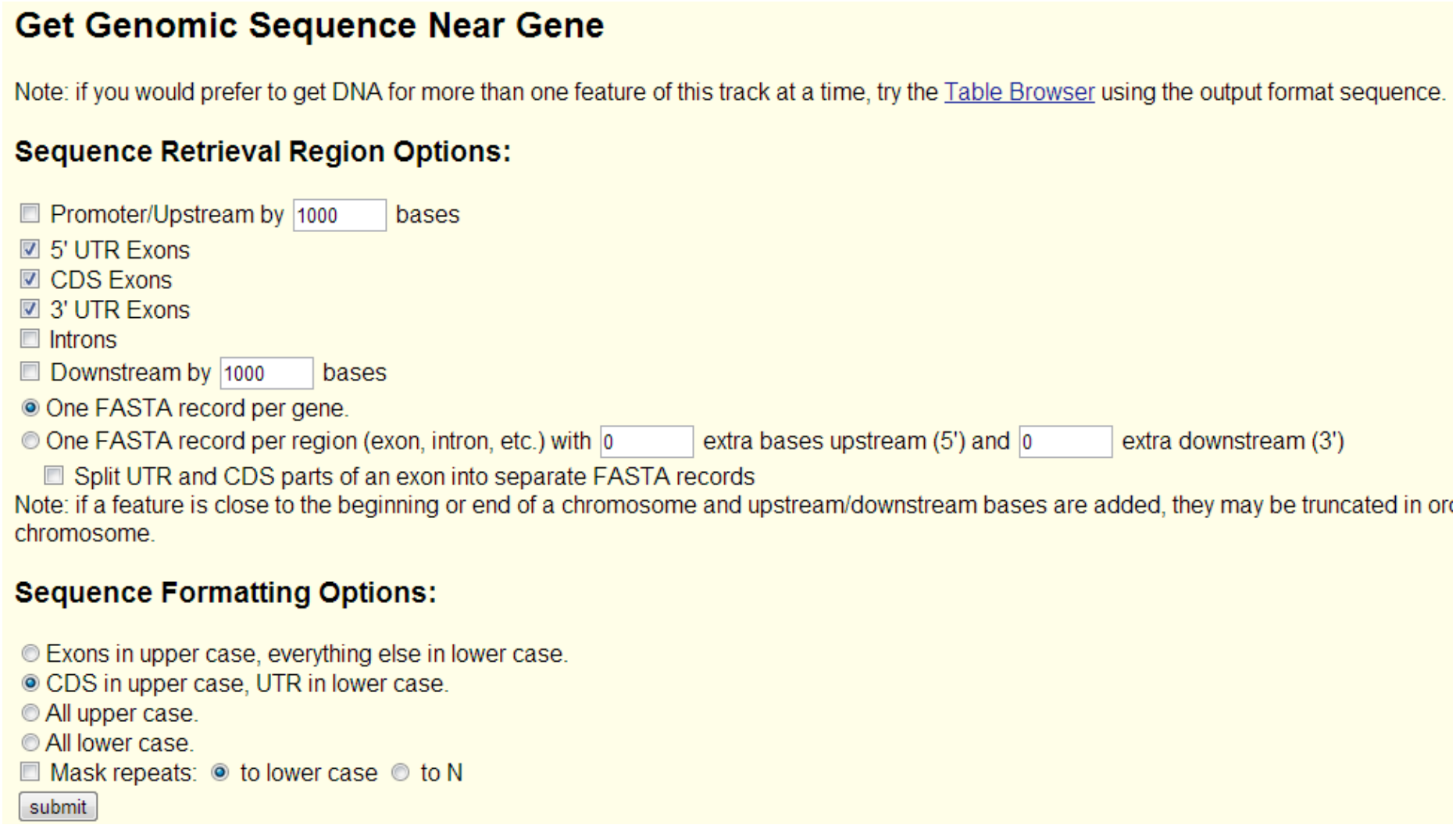
Go to RefSeq Genes track controls

Data last updated: 2013-01-25

Choose
Genomic sequence

MIRCORE

Let’s differentiate UTR and CDS



Get Genomic Sequence Near Gene

Note: if you would prefer to get DNA for more than one feature of this track at a time, try the Table Browser using the output format sequence.

Sequence Retrieval Region Options:

- ☐ Promoter/Upstream by 1000 bases
- ☑ 5' UTR Exons
- ☑ CDS Exons
- ☑ 3' UTR Exons
- ☐ Introns
- ☐ Downstream by 1000 bases
- ◉ One FASTA record per gene.
- ○ One FASTA record per region (exon, intron, etc.) with 0 extra bases upstream (5') and 0 extra downstream (3')
  - ☐ Split UTR and CDS parts of an exon into separate FASTA records

Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream bases are added, they may be truncated in or
chromosome.

Sequence Formatting Options:

- ○ Exons in upper case, everything else in lower case.
- ◉ CDS in upper case, UTR in lower case.
- ○ All upper case.
- ○ All lower case.
- ☐ Mask repeats: ◉ to lower case ○ to N

submit

MIRCORE

Output

```
>hg19_refGene_NM_000041 range=chr19:45409039-45412650
gggatccttgagtcctactcagccccagcggaggtgaaggacgtccttcc
ccaggagccgactggccaatcacaggcaggaagATGAAGGTTCTGTGGGC
TGCGTTGCTGGTCACATTCCTGGCAGGATGCCAGGCCAAGGTGGAGCAAG
CGGTGGAGACAGAGCCGGAGCCCGAGCTGCGCCAGCAGACCGAGTGGCAG
AGCGGCCAGCGCTGGGAACTGGCACTGGGTCGCTTTTGGGATTACCTGCG
CTGGGTGCAGACACTGTCTGAGCAGGTGCAGGAGGAGCTGCTCAGCTCCC
AGGTCACCCAGGAACTGAGGGCGCTGATGGACGAGACCATGAAGGAGTTG
AAGGCCTACAAATCGGAACTGGAGGAACAACTGACCCCGGTGGCGGAGGA
GACGCGGGCACGGCTGTCCAAGGAGCTGCAGGCGGCGCAGGCCCGGCTGG
GCGCGGACATGGAGGACGTGTGCGGCCGCCTGGTGCAGTACCGCGGCGAG
GTGCAGGCCATGCTCGGCCAGAGCACCGAGGAGCTGCGGGTGCGCCTCGC
CTCCCACCTGCGCAAGCTGCGTAAGCGGCTCCTCCGCGATGCCGATGACC
TGCAGAAGCGCCTGGCAGTGTACCAGGCCGGGGCCCGCGAGGGCGCCGAG
CGCGGCCTCAGCGCCATCCGCGAGCGCCTGGGGCCCCTGGTGGAACAGGG
CCGCGTGCGGGCCGCCACTGTGGGCTCCCTGGCCGGCCAGCCGCTACAGG
AGCGGGCCCAGGCCTGGGGCGAGCGGCTGCGCGCGCGGATGGAGGAGATG
GGCAGCCGGACCCGCGACCGCCTGGACGAGGTGAAGGAGCAGGTGGCGGA
GGTGCGCGCCAAGCTGGAGGAGCAGGCCCAGCAGATACGCCTGCAGGCCG
AGGCCTTCCAGGCCCGCCTCAAGAGCTGGTTCGAGCCCCTGGTGGAAGAC
ATGCAGCGCCAGTGGGCCGGGCTGGTGGAGAAGGTGCAGGCTGCCGTGGG
CACCAGCGCCGCCCCTGTGCCCAGCGACAATCACTGAacgccgaagcctg
cagccatgcgaccccacgccaccccgtgcctcctgcctccgcgcagcctg
cagcgggagaccctgtccccgccccagccgtcctcctggggtggacccta
gtttaataaagattcaccaagtttcacgca
```


MIRCORE

Tasks for your gene

- Find chromosome positions (start and end)
- Find entire DNA sequence with exons in capitals and introns in lower case.
- Find entire mature mRNA sequence with coding regions in capitals and UTRs in lower case.
- Find only coding regions.
- Find amino acid
- Connect 11^{th} – 20^{th} amino acids to the corresponding RNA coding sequences

Database to use:

http://genome.ucsc.edu/cgi-bin/hgGateway

http://www.ncbi.nlm.nih.gov/gene

Tasks for your gene

- Find all alternative splicing isoforms among RefSeq genes
  - http://genome.ucsc.edu/cgi-bin/hgGateway
- What are most commonly used exons among the isoforms?
  - Ex) exon1, exon2, exon9
- What are other genes directly interacting with your gene?
  - http://string-db.org