CAMM535
Fundamentals of Biological Databases

Final Project Report

Alzheimer's Disease Database

Nilüfer Baldır
Melis Oktayoğlu

Contents

Introduction to Alzheimer’s Disease

Alzheimer's disease is a progressive neurodegenerative disorder that affects memory, thinking, and behavior. Although there are multiple contributing factors, it is mainly caused by the accumulation of β-amyloid (Aβ) and tau proteins in the brain, loss of gray matter, and neuronal death. Symptoms typically begin with mild memory loss and difficulty completing familiar tasks, and progress to severe memory loss, confusion, difficulty communicating, and changes in mood and behavior. The disease has a strong genetic component, yet environmental and lifestyle factors also play an important role in its development. There is currently no cure for Alzheimer's, and current treatments primarily focus on managing symptoms.

There are two main types of genetic factors that have been identified in Alzheimer's disease:

1. **Late-onset Alzheimer's disease (LOAD):** The majority of Alzheimer's cases are late-onset, which means symptoms typically appear after age 60. The risk for LOAD is associated with genetic variations in several genes, the most important being apolipoprotein E (APOE). The APOE gene comes in three different versions, or alleles, called E2, E3, and E4. The E4 allele is associated with an increased risk of developing LOAD.
2. **Early-onset Alzheimer's disease (EOAD):** EOAD is a rarer form of Alzheimer's disease that occurs before age 60. EOAD is caused by mutations in specific genes such as presenilin 1 (PSEN1), presenilin 2 (PSEN2), and amyloid precursor protein (APP).

The biological processes leading to Alzheimer's disease are not fully understood, but research has identified several key mechanisms contributing to the progression of the disease.

1. Amyloid Beta (Aβ) accumulation: One of the hallmarks of Alzheimer's disease is the accumulation of Aβ peptides in the brain, which form plaques that are toxic to nerve cells. They are thought to disrupt communication between neurons and lead to the death of nerve cells.
2. Tau protein tangles: Another highly established contributor to Alzheimer's disease is the formation of tau protein tangles inside neurons. These tangles prevent the transport of nutrients and other important molecules within the cell, leading to the death of nerve cells.
3. Inflammation: Alzheimer's disease is also associated with chronic inflammation in the brain, which is thought to cause the death of nerve cells.

These are some of the main biological processes that are thought to contribute to the development of Alzheimer's disease, but the disease is still not fully understood and ongoing research is needed to gain more insight into the disease. (Ertekin-Taner, N, 2007)

Database Conceptual Design

For designing the database we prepared an ER diagram. The first table we constructed was the Var_Gene table which has the variants of Alzheimer's Disease with its associated genes. The associated genes is a multivalued attribute as a single variant can be in multiple genes. Other than that the other attributes we had in this table were phenotype description and name as we wanted to visualize how the disease is annotated and if our phenotype filter works properly as a form of sanity check.

The second table we constructed was the Genes_to_Coord table which has the primary key as the gene refSeq ID and holds the primary gene name, chromosome number, start and end location of the gene, as well as the exon count. We related this table to the Var_gene in a one-to-one relationship named "located in".

The third table we constructed was the UniProt table with the primary key UniProt ID, and the foreign key as the gene_name with reference to Genes_to_Coord. It had other attributes such as the protein name, mass, and length, multivalued attribute PDB ID as a protein can have many depositions to PDB, and primary gene name. For this table we created a relationship named "expressed" between it and the Genes_to_Coord which allows us to map a gene with its coordinates, to its structure related attributes.

The last table we had was the GEO2R table with Var_Gene table on their gene names with a relation called "studied gene". This table has the primary key as geo_id which is the id assigned by geo to the sample. Its other attributes are p value, adjusted p value, B and t values, logFC and gene symbol which might be an empty string.

Figure 1: *ER diagram of the Alzheimer database*

Physical Construction

1-Data Collection

a. Mapping Short Variant to Genes from BioMart

For the retrieval of the short variants and indels of Alzheimer's disease, we used BioMart. We selected the dataset Human Short Variants and applied the filter in the phenotype section as a single entry "Alzheimer disease". We selected the attributes to collect as the variant name, associated gene with phenotype, and phenotype name and description. As a result, we had 314 entries, meaning there are 314 SNPs and indels related to Alzheimer's. With a bash command, we parsed the unique names of genes in this table, which resulted in a total of 138 genes.

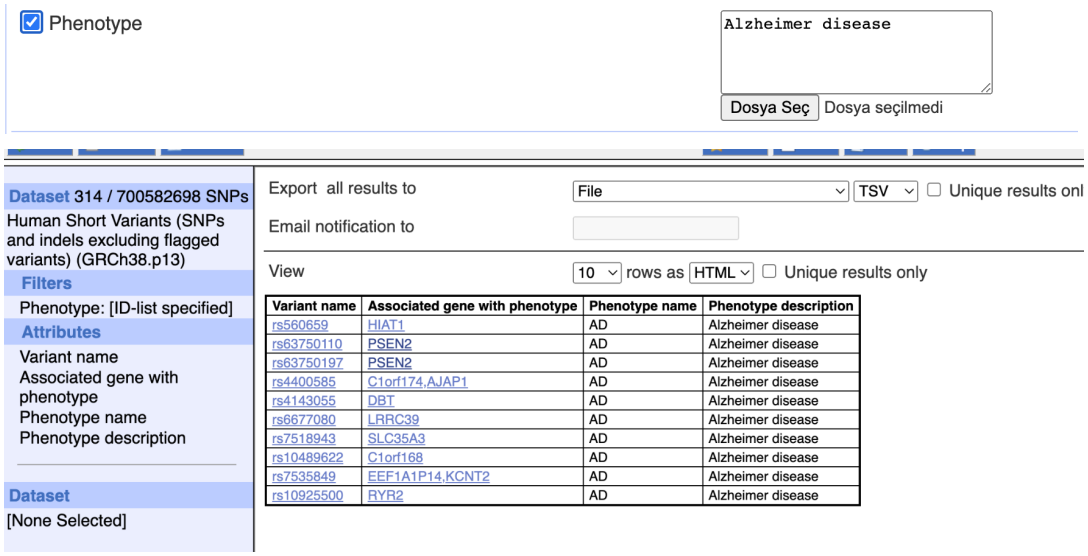


Figure 2: *On top, the filtering of the phenotype on BioMart, below the output of the query as well as the selected database and attributes on the right from the BioMart Website.*

b. Genomic Coordination Retrieval from USCS

Next, using these parsed unique gene names we got to the UCSC table browser, and by choosing from the knownGene table as can be seen below figure, we retrieved the genome coordinates after pasting the gene names to the identifiers list.

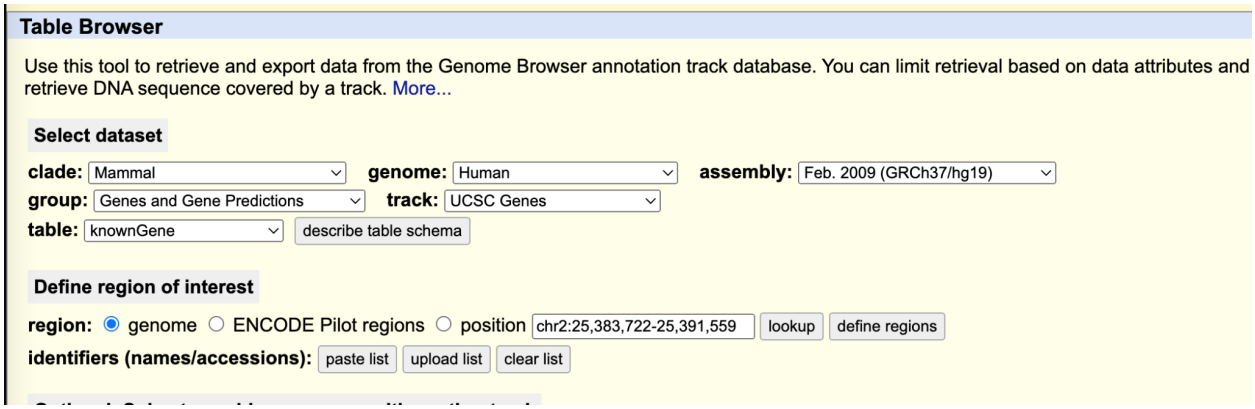


Figure 3: *Genomic location retrieval from UCSC table browser, using UCSC genes track and knownGene table.*

However, we had 21 of the gene names with a mapping error, thus we were left with 117 genes.

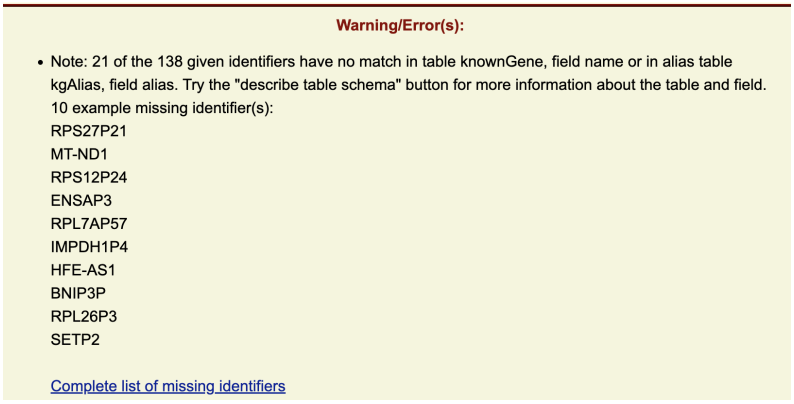


Figure 4: *Screenshot of the gene name mapping error we got for the 21 gene names out of 138 on the knownGene table.*

Below are the first 10 entries of the results we got, we intersected our query with the kgXref table as well to retrieve the geneSymbols and the refSeq IDs.

	A	B	C	D	E	F	G
1	#hg19.knownGene.name	hg19.knownGene.chrom	hg19.knownGene.txStart	hg19.knownGene.txEnd	hg19.knownGene.exonCount	hg19.kgXref.geneSymbol	hg19.kgXref.refseq
2	uc003dmg.3	chr3	64501330	64673365	40	ADAMTS9	NM_182920
3	uc004cyx.3	chrX	19007424	19140755	29	ADGRG2	NM_001079858
4	uc002bmt.2	chr15	89164526	89175512	4	AEN	NM_022767
5	uc001aln.3	chr1	4715104	4843851	6	AJAP1	NM_018836
6	uc004epr.3	chrX	112018104	112066372	11	AMOT	NM_001113490
7	uc003xom.3	chr8	41510743	41754280	43	ANK1	NM_001142446
8	uc002xze.1	chr20	57034425	57089949	4	APCDD1L	NM_153360
9	uc002pab.3	chr19	45409038	45412650	4	APOE	NM_000041
10	uc002ylz.3	chr21	27252860	27543138	18	APP	NM_000484

Figure 5: *First 10 entries of the Genes_to_coords table, the attributes are knonwGene.name as the primary key, chromosome number, start and end locations of the genes, exon counts, gene names and refSeqIDs.*

c. GEO2R Analysis

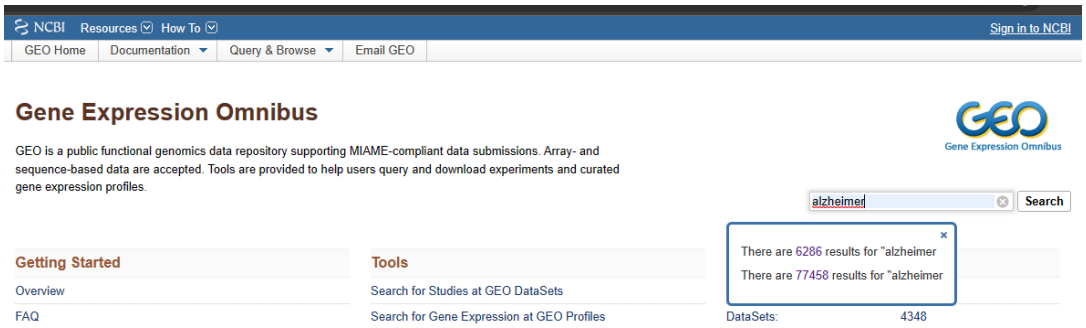
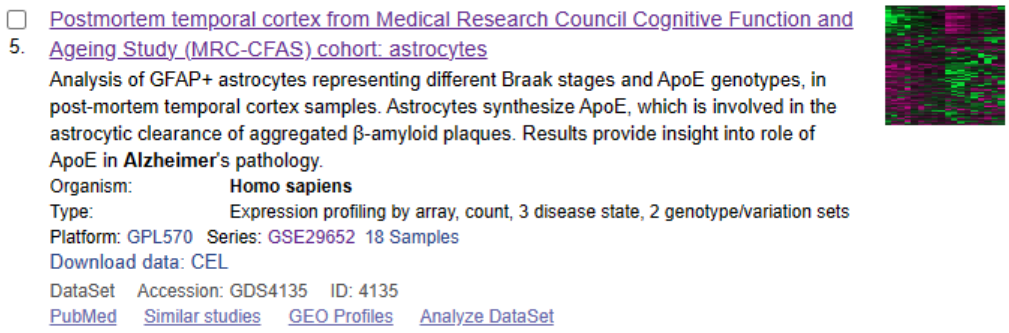


Figure 6: *Search for Alzheimer's Diseases in GEO database*

We searched Alzheimer's disease related datasets in GEO, and we choose the GSE29652 dataset.



5. Postmortem temporal cortex from Medical Research Council Cognitive Function and Ageing Study (MRC-CFAS) cohort: astrocytes

Analysis of GFAP+ astrocytes representing different Braak stages and ApoE genotypes, in post-mortem temporal cortex samples. Astrocytes synthesize ApoE, which is involved in the astrocytic clearance of aggregated β-amyloid plaques. Results provide insight into role of ApoE in **Alzheimer**'s pathology.

Organism: **Homo sapiens**

Type: Expression profiling by array, count, 3 disease state, 2 genotype/variation sets

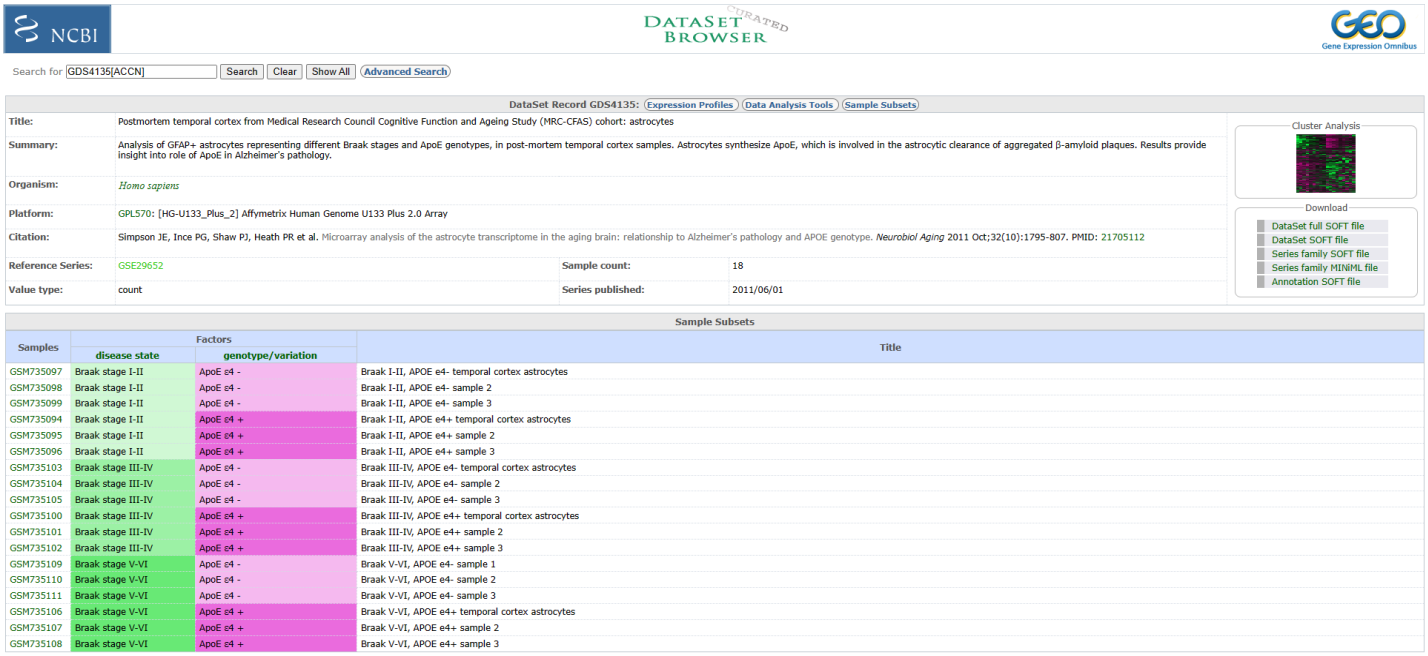
Platform: GPL570 Series: GSE29652 18 Samples

Download data: CEL

DataSet Accession: GDS4135 ID: 4135

PubMed Similar studies GEO Profiles Analyze DataSet

Figure 7: *Dataset for GEO2R analysis*



NCBI DATASET CURATED BROWSER GEO Gene Expression Omnibus

Search for GDS4135[ACCN] Search Clear Show All Advanced Search

DataSet Record GDS4135: Expression Profiles Data Analysis Tools Sample Subsets

Title:	Postmortem temporal cortex from Medical Research Council Cognitive Function and Ageing Study (MRC-CFAS) cohort: astrocytes
Summary:	Analysis of GFAP+ astrocytes representing different Braak stages and ApoE genotypes, in post-mortem temporal cortex samples. Astrocytes synthesize ApoE, which is involved in the astrocytic clearance of aggregated β-amyloid plaques. Results provide insight into role of ApoE in Alzheimer's pathology.
Organism:	*Homo sapiens*
Platform:	GPL570: [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array
Citation:	Simpson JE, Ince PG, Shaw PJ, Heath PR et al. Microarray analysis of the astrocyte transcriptome in the aging brain: relationship to Alzheimer's pathology and APOE genotype. *Neurobiol Aging* 2011 Oct;32(10):1795-807. PMID: 21705112
Reference Series:	GSE29652
Sample count:	18
Value type:	count
Series published:	2011/06/01

Cluster Analysis

Download

DataSet full SOFT file
DataSet SOFT file
Series family SOFT file
Series family MINiML file
Annotation SOFT file

Sample Subsets

Samples	Factors: disease state	Factors: genotype/variation	Title
GSM735097	Braak stage I-II	ApoE ε4 -	Braak I-II, APOE e4- temporal cortex astrocytes
GSM735098	Braak stage I-II	ApoE ε4 -	Braak I-II, APOE e4- sample 2
GSM735099	Braak stage I-II	ApoE ε4 -	Braak I-II, APOE e4- sample 3
GSM735094	Braak stage I-II	ApoE ε4 +	Braak I-II, APOE e4+ temporal cortex astrocytes
GSM735095	Braak stage I-II	ApoE ε4 +	Braak I-II, APOE e4+ sample 2
GSM735096	Braak stage I-II	ApoE ε4 +	Braak I-II, APOE e4+ sample 3
GSM735103	Braak stage III-IV	ApoE ε4 -	Braak III-IV, APOE e4- temporal cortex astrocytes
GSM735104	Braak stage III-IV	ApoE ε4 -	Braak III-IV, APOE e4- sample 2
GSM735105	Braak stage III-IV	ApoE ε4 -	Braak III-IV, APOE e4- sample 3
GSM735100	Braak stage III-IV	ApoE ε4 +	Braak III-IV, APOE e4+ temporal cortex astrocytes
GSM735101	Braak stage III-IV	ApoE ε4 +	Braak III-IV, APOE e4+ sample 2
GSM735102	Braak stage III-IV	ApoE ε4 +	Braak III-IV, APOE e4+ sample 3
GSM735109	Braak stage V-VI	ApoE ε4 -	Braak V-VI, APOE e4- sample 1
GSM735110	Braak stage V-VI	ApoE ε4 -	Braak V-VI, APOE e4- sample 2
GSM735111	Braak stage V-VI	ApoE ε4 -	Braak V-VI, APOE e4- sample 3
GSM735106	Braak stage V-VI	ApoE ε4 +	Braak V-VI, APOE e4+ temporal cortex astrocytes
GSM735107	Braak stage V-VI	ApoE ε4 +	Braak V-VI, APOE e4+ sample 2
GSM735108	Braak stage V-VI	ApoE ε4 +	Braak V-VI, APOE e4+ sample 3

Figure 8: *Sample Subset of the GSE29652*

The dataset contain 18 different sample; categorized 3 different stages (entorhinal stages (Braak stages 0–II), limbic stages (Braak Stages III–IV), and isocortical stages (Braak stages V–VI)). Within each category, 3 cases were contain which carried at least 1 ε4 allele and 3 cases which were APOE4 negative.

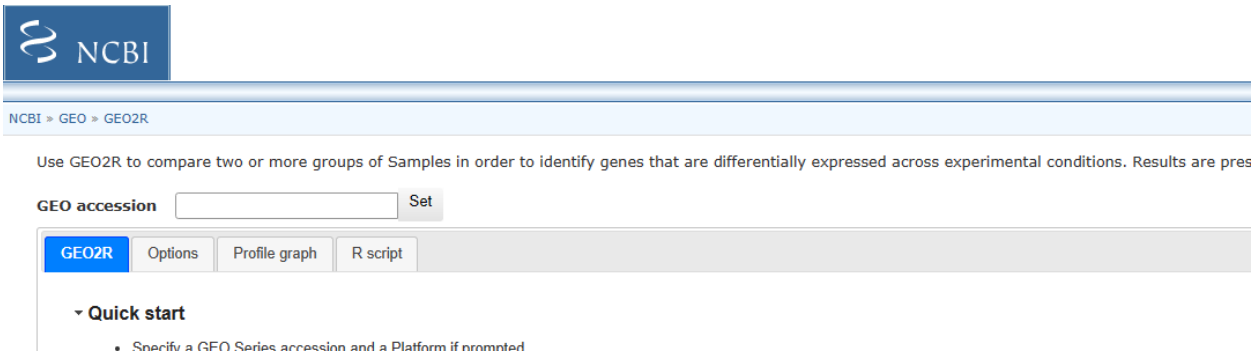
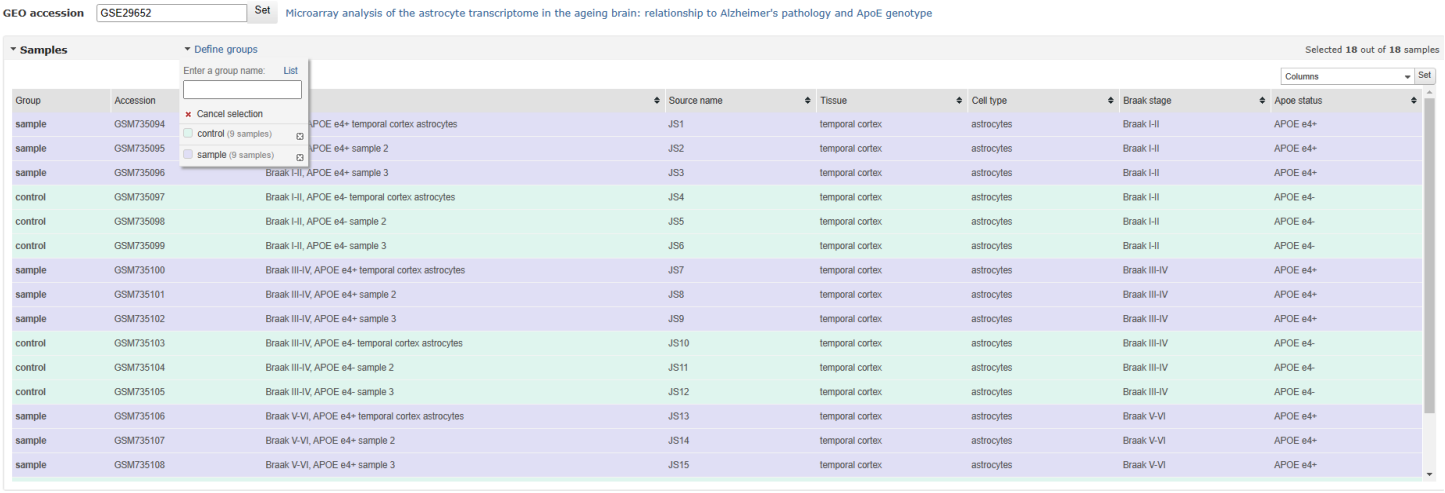


Figure 9: *GEO2R interface for GEO accession code*

We set the dataset in GEO2R.



Group	Accession		Source name	Tissue	Cell type	Braak stage	Apoe status
sample	GSM735094	APOE e4+ temporal cortex astrocytes	JS1	temporal cortex	astrocytes	Braak I-II	APOE e4+
sample	GSM735095	APOE e4+ sample 2	JS2	temporal cortex	astrocytes	Braak I-II	APOE e4+
sample	GSM735096	Braak I-II, APOE e4+ sample 3	JS3	temporal cortex	astrocytes	Braak I-II	APOE e4+
control	GSM735097	Braak I-II, APOE e4- temporal cortex astrocytes	JS4	temporal cortex	astrocytes	Braak I-II	APOE e4-
control	GSM735098	Braak I-II, APOE e4- sample 2	JS5	temporal cortex	astrocytes	Braak I-II	APOE e4-
control	GSM735099	Braak I-II, APOE e4- sample 3	JS6	temporal cortex	astrocytes	Braak I-II	APOE e4-
sample	GSM735100	Braak III-IV, APOE e4+ temporal cortex astrocytes	JS7	temporal cortex	astrocytes	Braak III-IV	APOE e4+
sample	GSM735101	Braak III-IV, APOE e4+ sample 2	JS8	temporal cortex	astrocytes	Braak III-IV	APOE e4+
sample	GSM735102	Braak III-IV, APOE e4+ sample 3	JS9	temporal cortex	astrocytes	Braak III-IV	APOE e4+
control	GSM735103	Braak III-IV, APOE e4- temporal cortex astrocytes	JS10	temporal cortex	astrocytes	Braak III-IV	APOE e4-
control	GSM735104	Braak III-IV, APOE e4- sample 2	JS11	temporal cortex	astrocytes	Braak III-IV	APOE e4-
control	GSM735105	Braak III-IV, APOE e4- sample 3	JS12	temporal cortex	astrocytes	Braak III-IV	APOE e4-
sample	GSM735106	Braak V-VI, APOE e4+ temporal cortex astrocytes	JS13	temporal cortex	astrocytes	Braak V-VI	APOE e4+
sample	GSM735107	Braak V-VI, APOE e4+ sample 2	JS14	temporal cortex	astrocytes	Braak V-VI	APOE e4+
sample	GSM735108	Braak V-VI, APOE e4+ sample 3	JS15	temporal cortex	astrocytes	Braak V-VI	APOE e4+

Figure 10: *Grouping the samples as sample vs control*

We defined two groups (APOE ε4+ (sample) and APOE ε4- (control)), and analyzed it. All results downloaded for phpmyadmin.

Top differentially expressed genes

Download full table Select columns

ID	adj.P.Val	P.Value	t	B	logFC	Gene.symbol	Gene.title
234582_at	0.0225	4.12e-07	-7.03	0.383	-3.46		
224515_at	0.1536	5.62e-06	-5.9	-0.381	-3.16		
1565702_at	0.2484	1.64e-05	-5.45	-0.729	-3.28	SMAD4	SMAD family member 4
234538_at	0.2484	1.82e-05	5.41	-0.763	3.46		
237253_at	0.5575	5.95e-05	4.92	-1.175	2.65	IGSF11-AS1	IGSF11 antisense RNA 1
1555853_at	0.5575	6.50e-05	-4.89	-1.206	-3.18	PSMB8-AS1	PSMB8 antisense RNA 1 (head to head)
216969_s_at	0.5575	7.65e-05	-4.82	-1.265	-2.73	KIF22	kinesin family member 22
224272_at	0.5575	8.16e-05	-4.79	-1.288	-3.79	RACGAP1P	Rac GTPase activating protein 1 pseud…
234608_at	0.5767	9.75e-05	-4.72	-1.353	-3.18	LAMA3	laminin subunit alpha 3
235280_at	0.5767	1.21e-04	-4.64	-1.432	-2.47	POLR1A	RNA polymerase I subunit A
241750_x_at	0.5767	1.30e-04	-4.61	-1.46	-3.58		
231211_s_at	0.5767	1.38e-04	4.58	-1.481	2.51	YIF1B	Yip1 interacting factor homolog B, mem…
241880_x_at	0.5767	1.42e-04	4.57	-1.493	3.44		
222134_at	0.5767	1.53e-04	-4.54	-1.519	-2.65	DDO	D-aspartate oxidase
236167_at	0.5767	1.58e-04	-4.53	-1.533	-2.8		
1570338_at	0.5957	1.93e-04	-4.45	-1.607	-2.54		

Figure 11: *GEO2R results*

d. UniProt: Mapping from Genes to Proteins

Finally, to retrieve the UniProt attributes we went to the UniProt website and through the ID mapping option we selected the Gene Name for the from option and the mapping target database we selected as the UniProtKB.

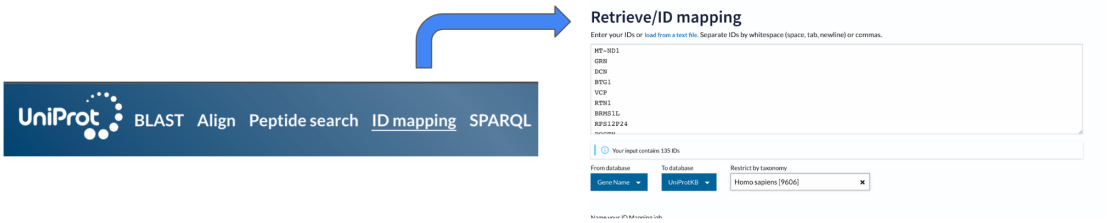


Figure 12: *A diagram describing the ID mapping step in UniProt. Gene names were mapped with UniProt IDs.*

This query resulted in 734 results, from which we selected the instructed attributes and downloaded the results in the form of a .tsv file. The attributes we collected were the gene names we used in mapping, UniProtIDs, full protein names, primary gene names, length, mass, and PDB ID of the protein entries. The genes have multiple protein entries as there can be multiple UniProt protein entries related to a single gene.

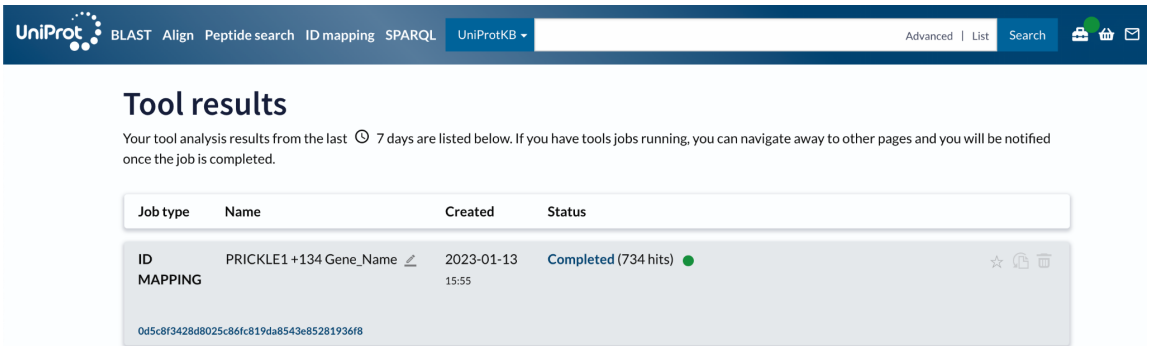


Figure 13: *Results from the ID mapping in UniProt, the search resulted in 734 found UniProt entires.*

	A	B	C	D	E	F	G
1	Gene_name	UniProt ID	Protein names	Gene Names	Length	Mass	PDB
2	PRICKLE1	Q96MT3	Prickle-like prote	PRICKLE1 RILP	831	94300	
3	PRICKLE1	A0A024R0W7	Prickle-like 1 (Dr	PRICKLE1 hCG_	831	94300	
4	PRICKLE1	A0A1W2PPC7	Prickle-like prote	PRICKLE1	128	14640	
5	PRICKLE1	F8VUG8	Prickle-like prote	PRICKLE1	163	18282	
6	PRICKLE1	F8W1J1	Prickle-like prote	PRICKLE1	119	13671	
7	PRICKLE1	F8W1Q8	Prickle-like prote	PRICKLE1	109	12588	
8	PTTG1	O95997	Securin (Esp1-a	PTTG1 EAP1 P	202	22024	7NJ0;7NJ1;
9	PTTG1	C4TNW4	Pituitary tumor-t	PTTG1	31	3488	
10	PTTG1	E5RJR4	Securin	PTTG1	121	12938	
11	PTTG1	Q2VPE7	PTTG1 protein	PTTG1	191	20826	
12	PTTG1	Q6IAL9	PTTG1 protein (I	PTTG1 hCG_16	202	22024	
13	ATP10B	O94823	Phospholipid-tra	ATP10B ATPVB	1461	165391	
14	ATP10B	A0A2R8YDI5	Phospholipid-tra	ATP10B hCG_1	1433	161612	
15	ATP10B	Q2YDW8	ATP10B protein	ATP10B	556	61764	

Figure 14: *First 15 results from the UniProt query.*

2- DB Construction on phpMyAdmin

For the deployment of all these data and to form the connections between the tables as we described in our ER diagram we went on the phpMyAdmin and opened a new database named "alzheimer_db".

Then, we created new tables with the same names as we assigned in the ER diagram, and specified the foreign key attributes as well as the attribute names.

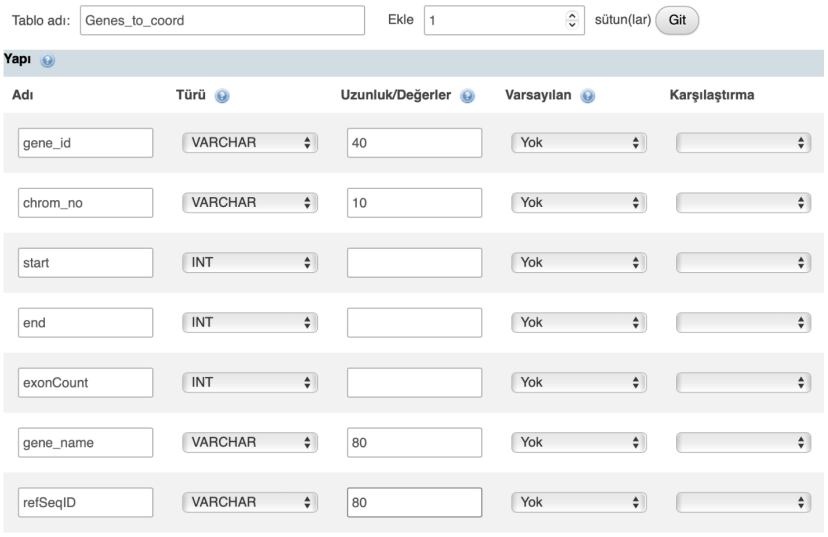


Figure 15: *The step in database construction in defining the tables, this example is from Genes_to_coord table.*

Genes_to_Coord

#	Adı	Türü	Karşılaştırma
1	gene_id	varchar(40)	utf8mb4_0900_ai_ci
2	chrom_no	varchar(10)	utf8mb4_0900_ai_ci
3	start	int	
4	end	int	
5	exonCount	int	
6	gene_name	varchar(80)	utf8mb4_0900_ai_ci
7	refSeqID	varchar(80)	utf8mb4_0900_ai_ci

GEO2R

#	Adı	Türü	Karşılaştırma
1	geo_id	varchar(40)	utf8mb4_0900_ai_ci
2	adj_p_val	double	
3	p_val	double	
4	t_val	double	
5	B_val	double	
6	logFC	double	
7	gene_symbol	varchar(500)	utf8mb4_0900_ai_ci

UniProt

#	Adı	Türü	Karşılaştırma
1	uniprot_id	varchar(80)	utf8mb4_0900_ai_ci
2	protein_name	varchar(1000)	utf8mb4_0900_ai_ci
3	primary_gene_name	varchar(400)	utf8mb4_0900_ai_ci
4	length	int	
5	mass	int	
6	PDB_id	varchar(1000)	utf8mb4_0900_ai_ci
7	gene_name	varchar(300)	utf8mb4_0900_ai_ci

Var_gene

#	Adı	Türü	Karşılaştırma
1	variant_name	varchar(80)	utf8mb4_0900_ai_ci
2	associated_gene	varchar(40)	utf8mb4_0900_ai_ci
3	pheno_desc	varchar(40)	utf8mb4_0900_ai_ci
4	pheno_name	varchar(40)	utf8mb4_0900_ai_ci

Figure 16: *A summary of the tables and its attributes.*

After constructing the tables we got the insertion SQL queries using a website called “ConvertSimple”. Where we paste the entries and the website creates the insertion query after we had to specify the exact table and attribute names before running on the phpMyAdmin.

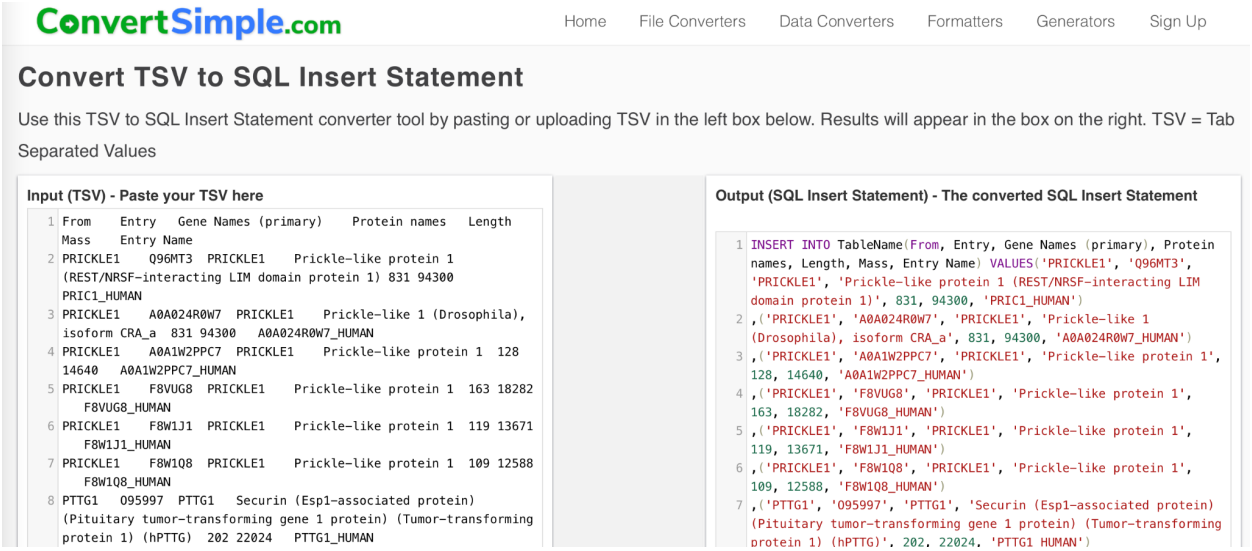


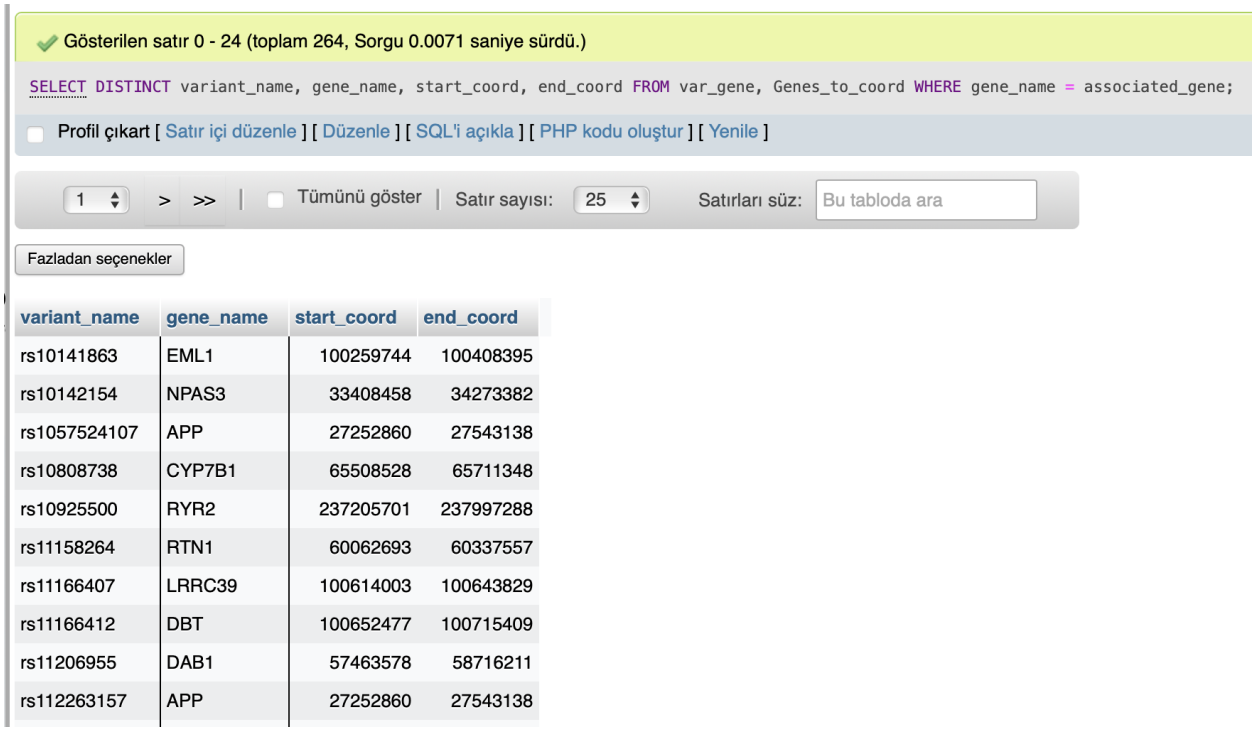
Figure 17: *Results from the ID mapping in UniProt, the search resulted in 734 found UniProt entries.*

Demo Queries

Q7- Write down at least 3 queries and SQL commands that will require at least 2 tables and include their results to demonstrate that your database has been constructed properly.

1- Our first query is for selecting coordinates of variants:

SELECT DISTINCT variant_name, gene_name, start_coord, end_coord
FROM var_gene, Genes_to_coord
WHERE gene_name = associated_gene;



variant_name	gene_name	start_coord	end_coord
rs10141863	EML1	100259744	100408395
rs10142154	NPAS3	33408458	34273382
rs1057524107	APP	27252860	27543138
rs10808738	CYP7B1	65508528	65711348
rs10925500	RYR2	237205701	237997288
rs11158264	RTN1	60062693	60337557
rs11166407	LRRC39	100614003	100643829
rs11166412	DBT	100652477	100715409
rs11206955	DAB1	57463578	58716211
rs112263157	APP	27252860	27543138

2- Our next query is for collecting gene names, logFC values, PDB IDs and chromosome coordinates of the proteins with length higher than 100, and which have a valid PDB entry.

SELECT DISTINCT gene_symbol, logFC, length, PDB_id, start_coord, end_coord

FROM GEO2R, UniProt, Genes_to_coord

WHERE length>100 AND UniProt.gene_name = GEO2R.gene_symbol AND
Genes_to_coord.gene_name = UniProt.gene_name AND PDB_id not like "";

Gösterilen satır 0 - 24 (toplam 147, Sorgu 0.0262 saniye sürdü.)

```
SELECT DISTINCT gene_symbol, logFC, length, PDB_id, start_coord, end_coord FROM GEO2R, UniProt, Genes_to_coord WHERE length>100 AND UniProt.gene_name = GEO2R.gene_symbol
AND Genes_to_coord.gene_name = UniProt.gene_name AND PDB_id not like "";
```

Profil çıkart [Satır içi düzenle] [Düzenle] [SQL'i açıkla] [PHP kodu oluştur] [Yenile]

1 > >> | Tümünü göster | Satır sayısı: 25 Satırları süz: Bu tabloda ara

Fazladan seçenekler

gene_symbol	logFC	length	PDB_id	start_coord	end_coord
APP	-0.52234721	485	6HAR;	27252860	27543138
APP	0.89379721	485	6HAR;	27252860	27543138
APP	-0.28788776	485	6HAR;	27252860	27543138
EML1	-0.04371091	815	4CI8;	100259744	100408395
EML1	-1.04188444	815	4CI8;	100259744	100408395
SPTLC1	-0.44033843	473	6M4N;6M4O;7CQI;7CQK;7K0I;7K0J;7K0K;7K0L;7K0M;7K0N;...	94793426	94877690
SPTLC1	-0.32961639	473	6M4N;6M4O;7CQI;7CQK;7K0I;7K0J;7K0K;7K0L;7K0M;7K0N;...	94793426	94877690
SPTLC1	-0.51186898	473	6M4N;6M4O;7CQI;7CQK;7K0I;7K0J;7K0K;7K0L;7K0M;7K0N;...	94793426	94877690
COBL	0.66370695	1261	4PL8;	51083908	51384515
COBL	-0.43524949	1261	4PL8;	51083908	51384515
TRPA1	0.16881066	1119	3J9P;6HC8;6PQO;6PQP;6PQQ;6V9V;6V9W;6V9X;6V9Y;6WJ5;...	72933485	72987819

3- Our next query is for selecting variants on chromosome 1 with expressed protein length smaller than 100.

SELECT variant_name, chrom_no, length, uniprot_id
FROM var_gene, Genes_to_coord, UniProt
WHERE var_gene.associated_gene = Genes_to_coord.gene_name AND Genes_to_coord.gene_name = UniProt.gene_name AND length <100 AND chrom_no like "chr1";

Gösterilen satır 0 - 15 (toplam 16, Sorgu 0.0046 saniye sürdü.)

```
SELECT variant_name, chrom_no, length, uniprot_id FROM var_gene, Genes_to_coord, UniProt WHERE var_gene.associated_gene = Genes_to_coord.gene_name AND
Genes_to_coord.gene_name = UniProt.gene_name AND length <100 AND chrom_no like "chr1";
```

Profil çıkart [Satır içi düzenle] [Düzenle] [SQL'i açıkla] [PHP kodu oluştur] [Yenile]

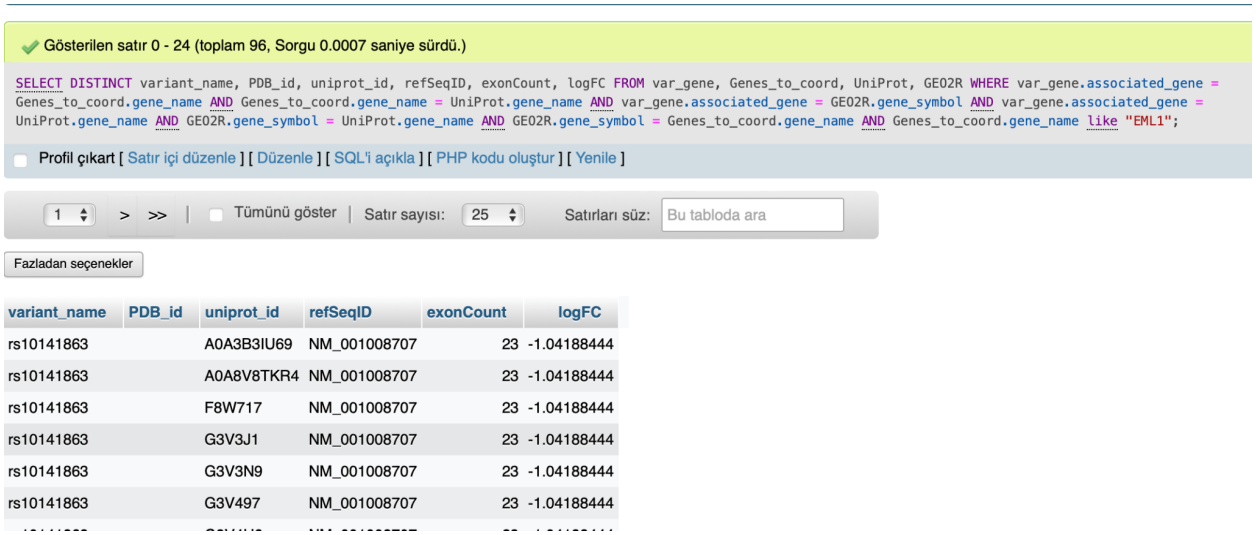
Tümünü göster | Satır sayısı: 25 Satırları süz: Bu tabloda ara

Fazladan seçenekler

variant_name	chrom_no	length	uniprot_id
rs10925500	chr1	34	A0A590UK06
rs10925500	chr1	54	L8E9M2
rs58973334	chr1	53	A0A7I2V5Y0
rs58973334	chr1	54	A0A7I2YQG9
rs58973334	chr1	21	E5RJM5
rs63750110	chr1	53	A0A7I2V5Y0
rs63750110	chr1	54	A0A7I2YQG9
rs63750110	chr1	21	E5RJM5
rs63750197	chr1	53	A0A7I2V5Y0
rs63750197	chr1	54	A0A7I2YQG9

Q8 - SQL command that will give an output of variation name, PDB IDs, Uniprot/SwissProt IDs, RefSeq IDs, exon counts, logFC value of a given gene symbol.

SELECT DISTINCT variant_name, PDB_id, uniprot_id, refSeqID, exonCount, logFC FROM var_gene, Genes_to_coord, UniProt, GEO2R WHERE var_gene.associated_gene = Genes_to_coord.gene_name AND Genes_to_coord.gene_name = UniProt.gene_name AND var_gene.associated_gene = GEO2R.gene_symbol AND var_gene.associated_gene = UniProt.gene_name AND GEO2R.gene_symbol = UniProt.gene_name AND GEO2R.gene_symbol = Genes_to_coord.gene_name AND Genes_to_coord.gene_name like "EML1";



variant_name	PDB_id	uniprot_id	refSeqID	exonCount	logFC
rs10141863		A0A3B3IU69	NM_001008707	23	-1.04188444
rs10141863		A0A8V8TKR4	NM_001008707	23	-1.04188444
rs10141863		F8W717	NM_001008707	23	-1.04188444
rs10141863		G3V3J1	NM_001008707	23	-1.04188444
rs10141863		G3V3N9	NM_001008707	23	-1.04188444
rs10141863		G3V497	NM_001008707	23	-1.04188444

Q9 - SQL query to retrieve a list of genes having PDB IDs and exon count is greater than or equal to 4

SELECT Genes_to_coord.gene_name,PDB_id, exonCount
FROM Genes_to_coord, UniProt WHERE UniProt.gene_name = Genes_to_coord.gene_name AND PDB_id not like "" AND exonCount>4;

Gösterilen satır 0 - 24 (toplam 41, Sorgu 0.0102 saniye sürdü.)

```
SELECT Genes_to_coord.gene_name,PDB_id, exonCount FROM Genes_to_coord, UniProt WHERE UniProt.gene_name = Genes_to_coord.gene_name AND PDB_id not like "" AND exonCount>4;
```

Profil çıkart [Satır içi düzenle] [Düzenle] [SQL'i açıkla] [PHP kodu oluştur] [Yenile]

1 > >> | Tümünü göster | Satır sayısı: 25 Satırları süz: Bu tabloda ara

Fazladan seçenekler

gene_name	PDB_id	exonCount
DBT	1K8M;1K8O;1ZWV;2COO;3RNM;	11
RYR2	4JKQ;6Y4O;6Y4P;7KL5;7U9Q;7U9R;7U9T;7U9X;7U9Z;7UA1;...	105
PRKCQ	1XJD;2ENJ;2ENN;2ENZ;2JED;4Q9Z;4RA5;5F9E;	18
MKI67	1R21;2AFF;5J28;	15
MGMT	1EH6;1EH7;1EH8;1QNT;1T38;1T39;1YFH;	5
PHF21A	2PUY;2YQL;	18
MRPL11	3J7Y;3J9M;5OOL;5OOM;6I9R;6NU2;6NU3;6VLZ;6VMI;6ZM5;...	5
CHORDC1	2YRT;	11
POSTN	5WT7;5YJG;5YJH;	23
DIAPH3	5UWP;6X2Y;	28
SEC23A	2NUP;2NUT;2YRC;2YRD;3EFO;3EG9;3EGD;3EGX;5KYN;5KYU;...	20
PSEN1	2KR6;4UIS;5A63;5FN2;5FN3;5FN4;5FN5;6IDF;6IYC;6LQG;...	12

Q10 - SQL query to count a list of genes whose adjusted p-value in GEO2R is smaller than 0.05

SELECT COUNT(gene_symbol) FROM GEO2R
WHERE adj_p_val < 0.05;

SQL sorgunuz başarılı olarak çalıştırıldı.

```
SELECT COUNT(gene_symbol) FROM GEO2R WHERE adj_p_val < 0.05;
```

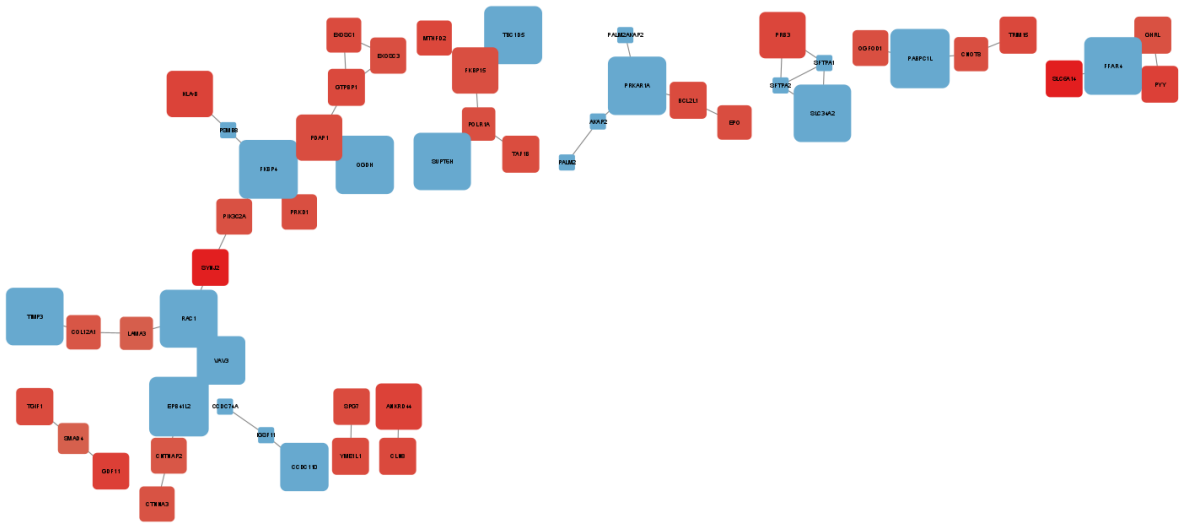
Profil çıkart [Satır içi düzenle] [Düzenle] [SQL'i açıkla] [PHP kodu oluştur] [Yenile]

Fazladan seçenekler

COUNT(gene_symbol)
1

From GEO dataset, we could not obtain significant adjusted p-value (lower than 0.05), so we used first 175 results of GEO2R analysis.

	A	B	C	D	E	F	G
1	ID	adj.P.Val	P.Value	t	B	logFC	Gene.symbol
2	234582_at	0.0225	4.12E-07	7.02759163	0.383	3.45564553	
3	224515_at	0.1536	5.62E-06	5.89743924	-0.381	3.16086967	
4	1565702_at	0.2484	1.64E-05	5.45076973	-0.729	3.28167683	SMAD4
5	234538_at	0.2484	1.82E-05	-5.40869639	-0.763	-3.46420309	
6	237253_at	0.5575	5.95E-05	-4.9228168	-1.175	-2.65076537	IGSF11-AS1
7	1555853_at	0.5575	6.50E-05	4.88692419	-1.206	3.17555374	PSMB8-AS1
8	216969_s_at	0.5575	7.65E-05	4.82104234	-1.265	2.73027656	KIF22
9	224272_at	0.5575	8.16E-05	4.79485412	-1.288	3.79405196	RACGAP1P
10	234608_at	0.5767	9.75E-05	4.72227702	-1.353	3.18437322	LAMA3
11	235280_at	0.5767	1.21E-04	4.63516116	-1.432	2.46932354	POLR1A



References:

Ertekin-Taner, N. (2007). Genetics of alzheimer's disease: A centennial review. *Neurologic Clinics*, *25*(3), 611–667. https://doi.org/10.1016/j.ncl.2007.03.009