

TERM PROJECT

BIN503: BIOLOGICAL DATABASES AND DATA ANALYSIS TOOLS

Asli YAZAGAN & Gungor BUDAK
June 11, 2013

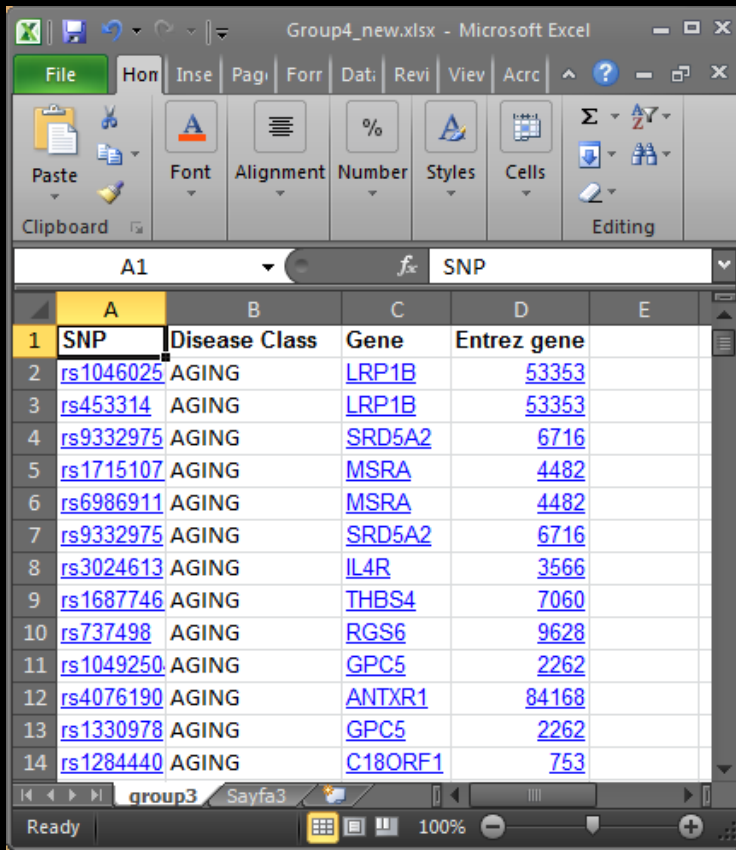
Project Description

- Developing an online (web-based) search tool on
 - a given database
 - a retrieved database from Biomart Ensembl
 - A list containing human genes without any known protein structure
 - Find mouse orthologs of genes on the list
 - Show Ensembl gene ID, chromosome number, position, HGNC symbol, human PDB ID, mouse PDB ID

Preparing Given Table

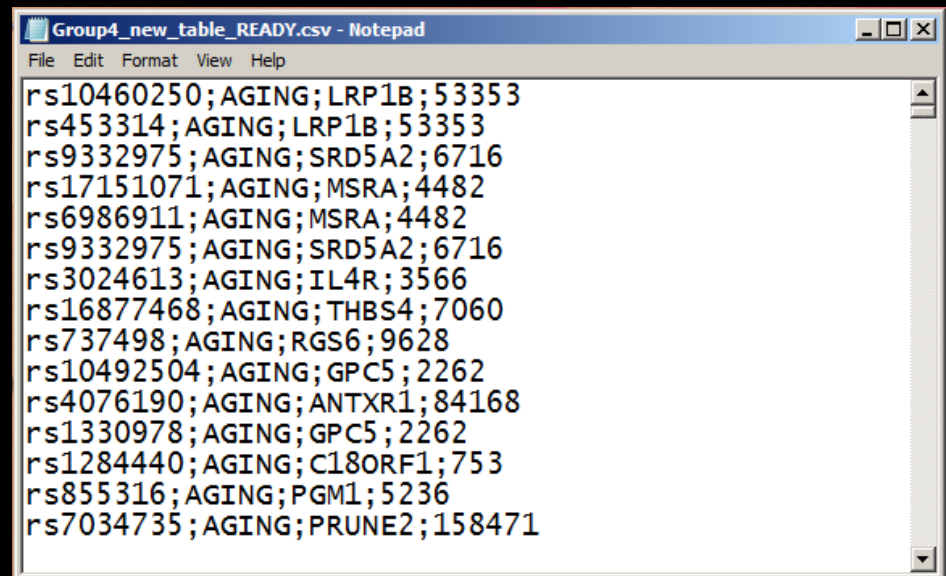
- XLS or XLSX format
 - Not suitable for insertion to MySQL database
 - Needs to be converted to CSV
- CSV format
 - Easily obtained by «Saving As» XLS or XLSX on MS

XLSX vs CSV



Group4_new.xlsx - Microsoft Excel

	A1		f _x	SNP	
	A	B	C	D	E
1	SNP	Disease Class	Gene	Entrez gene	
2	rs1046025	AGING	LRP1B	53353	
3	rs453314	AGING	LRP1B	53353	
4	rs9332975	AGING	SRD5A2	6716	
5	rs1715107	AGING	MSRA	4482	
6	rs6986911	AGING	MSRA	4482	
7	rs9332975	AGING	SRD5A2	6716	
8	rs3024613	AGING	IL4R	3566	
9	rs1687746	AGING	THBS4	7060	
10	rs737498	AGING	RGS6	9628	
11	rs1049250	AGING	GPC5	2262	
12	rs4076190	AGING	ANTXR1	84168	
13	rs1330978	AGING	GPC5	2262	
14	rs1284440	AGING	C18ORF1	753	



Group4_new_table_READY.csv - Notepad

```
File Edit Format View Help
rs1046025;AGING;LRP1B;53353
rs453314;AGING;LRP1B;53353
rs9332975;AGING;SRD5A2;6716
rs1715107;AGING;MSRA;4482
rs6986911;AGING;MSRA;4482
rs9332975;AGING;SRD5A2;6716
rs3024613;AGING;IL4R;3566
rs1687746;AGING;THBS4;7060
rs737498;AGING;RGS6;9628
rs1049250;AGING;GPC5;2262
rs4076190;AGING;ANTXR1;84168
rs1330978;AGING;GPC5;2262
rs1284440;AGING;C18ORF1;753
rs855316;AGING;PGM1;5236
rs7034735;AGING;PRUNE2;158471
```

MySQL DB Table Structures

```
SELECT *  
FROM `given`  
LIMIT 0 , 30
```

☐ Profil çıkart [Sıralı] [Düzenle] [SQL'i açıkla] [PHP Kodu oluştur] [Yenile]

#	Sütun	Türü	Karşılaştırma	Öznitelikler	Boş	Varsayılan	Ekstra	Eylem
<input type="checkbox"/>	1 rsid	varchar(255)	latin1_swedish_ci		Hayır	Yok		Değiştir Kaldır Daha Fazla ▼
<input type="checkbox"/>	2 disease_class	varchar(255)	latin1_swedish_ci		Hayır	Yok		Değiştir Kaldır Daha Fazla ▼
<input type="checkbox"/>	3 hgnc_symbol	varchar(255)	latin1_swedish_ci		Hayır	Yok		Değiştir Kaldır Daha Fazla ▼
<input type="checkbox"/>	4 entrez_gene_id	varchar(255)	latin1_swedish_ci		Hayır	Yok		Değiştir Kaldır Daha Fazla ▼

↑ Tümünü Seç / Hiçbirini Seçme Seçilileri: Gözet Değiştir Kaldır Birincil Benzersiz

```
SELECT *  
FROM `created`  
LIMIT 0 , 30
```

☐ Profil çıkart [Sıralı] [Düzenle] [SQL'i açıkla] [PHP Kodu oluştur] [Yenile]

#	Sütun	Türü	Karşılaştırma	Öznitelikler	Boş	Varsayılan	Ekstra	Eylem
<input type="checkbox"/>	1 human_ensembl_gene_id	varchar(255)	latin1_swedish_ci		Hayır	Yok		Değiştir Kaldır Daha Fazla ▼
<input type="checkbox"/>	2 chromosome_number	varchar(255)	latin1_swedish_ci		Hayır	Yok		Değiştir Kaldır Daha Fazla ▼
<input type="checkbox"/>	3 gene_start	varchar(255)	latin1_swedish_ci		Hayır	Yok		Değiştir Kaldır Daha Fazla ▼
<input type="checkbox"/>	4 gene_end	varchar(255)	latin1_swedish_ci		Hayır	Yok		Değiştir Kaldır Daha Fazla ▼
<input type="checkbox"/>	5 human_pdb_id	varchar(255)	latin1_swedish_ci		Hayır	Yok		Değiştir Kaldır Daha Fazla ▼
<input type="checkbox"/>	6 hgnc_symbol	varchar(255)	latin1_swedish_ci		Hayır	Yok		Değiştir Kaldır Daha Fazla ▼
<input type="checkbox"/>	7 mouse_ensembl_gene_id	varchar(255)	latin1_swedish_ci		Hayır	Yok		Değiştir Kaldır Daha Fazla ▼
<input type="checkbox"/>	8 mouse_pdb_id	varchar(255)	latin1_swedish_ci		Hayır	Yok		Değiştir Kaldır Daha Fazla ▼

↑ Tümünü Seç / Hiçbirini Seçme Seçilileri: Gözet Değiştir Kaldır Birincil Benzersiz İndeks

Insertion of CSV to DB

VERİ YÜKLE kullanarak CSV

BİÇİM-BELİRLİ SEÇENEKLER:

☐ Tablo versiyonunu dosya ile değiştir

☐ EKLEME hatasında durdurma

Sütunu sonlandırıcı:

Sütunu kapsayan:

Sütunu dolduran:

Satır sonlandırıcı:

Sütun adları:

☒ YEREL anahtar kelime kullan

Göt

localhost ▶ 503projectdb ▶ given

Gözet Yapı SQL Ara Ekle Dışa Aktar İçe Aktar İşlemler
























✓ İçe aktarma başarılı olarak bitti, 1 sorgu çalıştırıldı. (Group4_new_table_READY.csv)

LOAD DATA LOCAL INFILE 'C:\\wamp\\tmp\\phpF7FC.tmp' INTO TABLE `given` FIELDS TERMINATED BY ',' LINES TERMINATED BY '\r\n' # 3292 satır etkilendi.

[Sıralı] [Düzenle] [PHP Kodu oluştur]

Insertion of Given Table

+ Seçenekler

					rsid	disease_class	hgnc_symbol	entrez_gene_id				
<input type="checkbox"/>		Düzenle		Sıralı Düzenleme		Kopyala		Sil	rs10460250	AGING	LRP1B	53353
<input type="checkbox"/>		Düzenle		Sıralı Düzenleme		Kopyala		Sil	rs453314	AGING	LRP1B	53353
<input type="checkbox"/>		Düzenle		Sıralı Düzenleme		Kopyala		Sil	rs9332975	AGING	SRD5A2	6716
<input type="checkbox"/>		Düzenle		Sıralı Düzenleme		Kopyala		Sil	rs17151071	AGING	MSRA	4482
<input type="checkbox"/>		Düzenle		Sıralı Düzenleme		Kopyala		Sil	rs6986911	AGING	MSRA	4482
<input type="checkbox"/>		Düzenle		Sıralı Düzenleme		Kopyala		Sil	rs9332975	AGING	SRD5A2	6716
<input type="checkbox"/>		Düzenle		Sıralı Düzenleme		Kopyala		Sil	rs3024613	AGING	IL4R	3566
<input type="checkbox"/>		Düzenle		Sıralı Düzenleme		Kopyala		Sil	rs16877468	AGING	THBS4	7060
<input type="checkbox"/>		Düzenle		Sıralı Düzenleme		Kopyala		Sil	rs737498	AGING	RGS6	9628
<input type="checkbox"/>		Düzenle		Sıralı Düzenleme		Kopyala		Sil	rs10492504	AGING	GPC5	2262
<input type="checkbox"/>		Düzenle		Sıralı Düzenleme		Kopyala		Sil	rs4076190	AGING	ANTXR1	84168
<input type="checkbox"/>		Düzenle		Sıralı Düzenleme		Kopyala		Sil	rs1330978	AGING	GPC5	2262
<input type="checkbox"/>		Düzenle		Sıralı Düzenleme		Kopyala		Sil	rs1284440	AGING	C18ORF1	753
<input type="checkbox"/>		Düzenle		Sıralı Düzenleme		Kopyala		Sil	rs855316	AGING	PGM1	5236
<input type="checkbox"/>		Düzenle		Sıralı Düzenleme		Kopyala		Sil	rs7034735	AGING	PRUNE2	158471
<input type="checkbox"/>		Düzenle		Sıralı Düzenleme		Kopyala		Sil	rs196328	AGING	BAG3	9531

Retrieving Table form Biomart

Dataset
Homo sapiens genes (GRCh37.p10)
Filters
HGNC symbol(s) [e.g. ZFY]: [ID-list specified]
<u>with</u> PDB ID(s): Excluded
Orthologous Mouse Genes: Only
Attributes
<u>Ensembl</u> Gene ID
Chromosome Name
Gene Start (<u>bp</u>)
Gene End (<u>bp</u>)
HGNC symbol
PDB ID

Dataset
<u>Mus musculus</u> genes (GRCm38.p1)
Filters
[None selected]
Attributes
<u>Ensembl</u> Gene ID
PDB ID

Search Results as XLS to DB

Export all results to ☒ Unique results only

Email notification to

View rows as ☒ Unique results only

Ensembl Gene ID	Chromosome Name	Gene Start (bp)	Gene End (bp)	HGNC symbol	PDB ID	Ensembl Gene ID	PDB ID
ENSG00000183715	11	132284871	133402414	OPCML		ENSMUSG00000062257	
ENSG00000155052	2	124782864	125672864	CNTNAP5		ENSMUSG00000067028	
ENSG00000155052	2	124782864	125672864	CNTNAP5		ENSMUSG00000070695	
ENSG00000179915	2	50145643	51259674	NRXN1		ENSMUSG00000024109	3BOD
ENSG00000179915	2	50145643	51259674	NRXN1		ENSMUSG00000024109	3MW2
ENSG00000179915	2	50145643	51259674	NRXN1		ENSMUSG00000024109	
ENSG00000183117	8	2792992	4852494	CSMD1		ENSMUSG00000060924	
ENSG00000164532	7	35242042	35293758	TBX20		ENSMUSG00000031965	
ENSG00000166105	11	134144139	134189458	GLB1L3		ENSMUSG00000031966	
ENSG00000149177	11	48002113	48189670	PTPRJ		ENSMUSG00000025314	

- The same method is used as before when the given table is inserted

Biomart Results Table




human_ensembl_gene_id	chromosome_number	gene_start	gene_end	human_pdb_id	hgnc_symbol	mouse_ensembl_gene_id	mouse_pdb_id
ENSG00000183715	11	132284871	133402414		OPCML	ENSMUSG00000062257	
ENSG00000155052	2	124782864	125672864		CNTNAP5	ENSMUSG00000067028	
ENSG00000155052	2	124782864	125672864		CNTNAP5	ENSMUSG00000070695	
ENSG00000179915	2	50145643	51259674		NRXN1	ENSMUSG00000024109	3BOD
ENSG00000179915	2	50145643	51259674		NRXN1	ENSMUSG00000024109	3MW2
ENSG00000179915	2	50145643	51259674		NRXN1	ENSMUSG00000024109	
ENSG00000183117	8	2792992	4852494		CSMD1	ENSMUSG00000060924	
ENSG00000164532	7	35242042	35293758		TBX20	ENSMUSG00000031965	
ENSG00000166105	11	134144139	134189458		GLB1L3	ENSMUSG00000031966	
ENSG00000149177	11	48002113	48189670		PTPRJ	ENSMUSG00000025314	
ENSG00000197584	3	177990720	178562217		KCNMB2	ENSMUSG00000037610	
ENSG00000104714	8	564746	688106		ERICH1	ENSMUSG000000051978	
ENSG00000151952	12	129556270	130388211		TMEM132D	ENSMUSG00000034310	
ENSG00000149571	11	126293254	126873355		KIRREL3	ENSMUSG00000032036	
ENSG00000154928	3	134316643	134979309		EPHB1	ENSMUSG00000032537	
ENSG00000139364	12	125671382	126146917		TMEM132B	ENSMUSG00000070498	

Search Tool Features

- PHP
- AJAX (for asynchronous calls)
- jQuery (JavaScript library to do AJAX calls and manipulate HTML objects)
- MySQL as DB
- CSS styling (to give a better look)
- UTF-8 encoding

Website Home

BIN503 Project - Windows Internet Explorer

   <http://bin503project.byethost18.com/>     BIN503 Project 

BIN503 Project

[Hide advanced search options](#)

☐ RSID ☐ Disease Class ☒ HGNC Symbol ☐ Entrez Gene ID

☐ Human Ensembl Gene ID ☐ Chromosome Number ☐ Gene Start (bp)

☐ Gene End (bp) ☐ Human PDB ID ☐ Mouse Ensembl Gene ID

☐ Mouse PDB ID

Yazagan & Budak

A Search on Website

mas

[Hide advanced search options](#)

- ☐ RSID ☒ Disease Class ☒ HGNC Symbol ☐ Entrez Gene ID
- ☐ Human Ensembl Gene ID ☒ Chromosome Number ☒ Gene Start (bp)
- ☒ Gene End (bp) ☐ Human PDB ID ☐ Mouse Ensembl Gene ID
- ☐ Mouse PDB ID

disease_class	hgnc_symbol	chromosome_number	gene_start	gene_end
CARDIOVASCULAR	MAST4	5	65892176	66465423
CARDIOVASCULAR	MAST4	5	65892176	66465423
CARDIOVASCULAR	MAST4	5	65892176	66465423
CARDIOVASCULAR	MAST4	5	65892176	66465423
CHEMDEPENDENCY	MAST4	5	65892176	66465423
CHEMDEPENDENCY	MAST4	5	65892176	66465423
CHEMDEPENDENCY	MAST4	5	65892176	66465423
CHEMDEPENDENCY	MAST4	5	65892176	66465423

[Clear results](#)

Live Demonstration

- [Project website](#)

Thanks for listening

Q/A