

Esercitazione UCSC – Table Browser.

Obiettivi dell'esercitazione:

- (I) estrarre una mappa di conversione dei principali identificativi (UCSC,Ensembl,HGNC,RefSeq) relativa a tutti i geni noti nella versione del genoma umano del febbraio 2009 (GRCh37/hg19).
- (II) importare e visualizzare il risultato ottenuto dal Table Browser in un foglio di calcolo

parte I:
estrarre una mappa di conversione dei principali identificativi (UCSC,Ensembl...) relativa a tutti i geni noti nella versione del genoma umano del febbraio 2009 (GRCh37/hg19).

1. apri una pagina del UCSC genome browser <http://genome.ucsc.edu>
2. dal menù principale --> Tools --> Table Browser
3. definisci i criteri di ricerca come segue:
clade:Mammal genome:Human assembly: Feb.2009(GRCh37/hg19)
group:Genes and Gene Predictions track: UCSC Genes
table:knownGene
region:genome
output format:selected fields from primary and related tables
output file:gene_ids_map__ucsc.hg19.txt

The screenshot shows the UCSC Genome Browser Table Browser interface. The browser's address bar displays genome.ucsc.edu/cgi-bin/hgTables. The page title is "Table Browser". Below the title, there is a detailed description of the tool's purpose: to retrieve data associated with a track in text format, calculate intersections between tracks, and retrieve DNA sequence covered by a track. It also provides links to the User's Guide, OpenHelix Table Browser tutorial, Galaxy, public MySQL server, GREAT, GenomeSpace, Credits, and Sequence and Annotation Downloads page.

The form contains the following fields and controls:

- clade:** Mammal (dropdown)
- genome:** Human (dropdown)
- assembly:** Feb. 2009 (GRCh37/hg19) (dropdown)
- group:** Genes and Gene Predictions (dropdown)
- track:** UCSC Genes (dropdown)
- add custom tracks** (button)
- track hubs** (button)
- table:** knownGene (dropdown)
- describe table schema** (button)
- region:** ☒ genome ☐ ENCODE Pilot regions ☐ position chr21:33,031,597-33,041,5
- lookup** (button)
- define regions** (button)
- identifiers (names/accessions):** **paste list** (button) **upload list** (button)
- filter:** **create** (button)
- intersection:** **create** (button)
- correlation:** **create** (button)
- output format:** selected fields from primary and related tables (dropdown)
- Send output to:** ☐ Galaxy ☐ GREAT ☐ GenomeSpace
- output file:** (text input field) (leave blank to keep output in browser)
- file type returned:** ☒ plain text ☐ gzip compressed
- get output** (button)
- summary/statistics** (button)

At the bottom, there is a link: "To reset all user cart settings (including custom tracks), [click here](#)."

4. premi "get output" --> verrai indirizzato ad una nuova pagina con la possibilità di selezionare vari campi -sia dalla tabella "knownGene" che da altre tabelle correlate - da riportare nel file di output;

4.a spunta le tabelle correlate elencate qui di seguito e poi premi il pulsante "allow selection from checked tables" che si trova in fondo alla pagina:

"hg19 – knownToEnsembl", "hg19 – ensGene", "hg19 – kgXref", "hg19 - knownToLocusLink"

4.b seleziona i seguenti campi dalla tabella "knownGene" e dalle tabelle correlate:

[dalla tabella "hg19.knownGene"] name;

[dalla tabella "hg19.ensGene"] name2;

[dalla tabella "hg19.kgXref"] geneSymbol e refseq;

[dalla tabella "hg19.knownToEnsembl"] value;

[dalla tabella "hg19.knownToLocusLink"] value;

Select Fields from hg19.knownGene - Chromium

genome.ucsc.edu/cgi-bin/hgTables

BIO_PROJECTS http://www.nat... BIO_Methods BIO_DBs NEWS BIO_PROJECTS temp_varie New folder SnowboardA... Other Bookmarks

Genomes Genome Browser Tools Mirrors Downloads My Data Help About Us

Select Fields from hg19.knownGene

<input type="checkbox"/>	name	Name of gene
<input type="checkbox"/>	chrom	Reference sequence chromosome or scaffold
<input type="checkbox"/>	strand	+ or - for strand
<input type="checkbox"/>	txStart	Transcription start position
<input type="checkbox"/>	txEnd	Transcription end position
<input type="checkbox"/>	cdsStart	Coding region start
<input type="checkbox"/>	cdsEnd	Coding region end
<input type="checkbox"/>	exonCount	Number of exons
<input type="checkbox"/>	exonStarts	Exon start positions
<input type="checkbox"/>	exonEnds	Exon end positions
<input type="checkbox"/>	proteinID	UniProt display ID for Known Genes, UniProt accession or RefSeq protein ID for UCSC Genes
<input type="checkbox"/>	alignID	Unique identifier for each (known gene, alignment position) pair

get output cancel check all clear all

hg19.kgXref fields

<input type="checkbox"/>	kgID	Known Gene ID
<input type="checkbox"/>	mRNA	mRNA ID
<input type="checkbox"/>	spID	UniProt protein Accession number
<input type="checkbox"/>	spDisplayID	UniProt display ID

4.c premi il pulsante "get output".

genome.ucsc.edu/cgi-bin/hgTables - Chromium

genome.ucsc.edu/cgi-bin/

genome.ucsc.edu/cgi-bin/hgTables

BIO_PROJECTS http://www.nat... BIO_Methods BIO_DBS NEWS BIO_PROJECTS temp_varie New folder SnowboardA... Other Bookmarks

hg19.knownGene.name	hg19.kgXref.kgID	hg19.kgXref.geneSymbol	hg19.kgXref.refseq	hg19.knownToEnsembl.value	hg19.knownToLocusLink.value	hg19.refGene.name2
uc001aaa.3	uc001aaa.3	DDX11L1 NR_046018	ENST00000456328	100287102	n/a	
uc010nxx.1	uc010nxx.1	DDX11L1	ENST00000456328	100287102	n/a	
uc010nxq.1	uc010nxq.1	DDX11L1	ENST00000518655	100287102	n/a	
uc009vis.3	uc009vis.3	WASH7P	ENST00000423562	653635	n/a	
uc009vjc.1	uc009vjc.1	WASH7P	ENST00000541675	653635	n/a	
uc009vjd.2	uc009vjd.2	WASH7P	ENST00000488147	653635	n/a	
uc009vit.3	uc009vit.3	WASH7P	ENST00000438504	653635	n/a	
uc009viu.3	uc009viu.3	WASH7P	ENST00000438504	653635	n/a	
uc001aae.4	uc001aae.4	WASH7P	ENST00000438504	653635	n/a	
uc001aai.1	uc001aai.1	WASH7P	ENST00000438504	653635	n/a	
uc001aah.4	uc001aah.4	WASH7P NR_024540	ENST00000438504	653635	n/a	
uc009vir.3	uc009vir.3	WASH7P	ENST00000438504	653635	n/a	
uc009viq.3	uc009viq.3	WASH7P	ENST00000423562	653635	n/a	
uc001aac.4	uc001aac.4	WASH7P	ENST00000438504	653635	n/a	
uc009viv.2	uc009viv.2	WASH7P	ENST00000423562	653635	n/a	
uc009viw.2	uc009viw.2	WASH7P	ENST00000438504	653635	WASH7P,	
uc009vix.2	uc009vix.2	WASH7P	ENST00000423562	653635	n/a	
uc009viy.2	uc009viy.2	WASH7P	ENST00000438504	653635	n/a	
uc009viz.2	uc009viz.2	WASH7P	ENST00000438504	653635	n/a	
uc010nxs.1	uc010nxs.1	WASH7P	ENST00000438504	653635	n/a	
uc009vje.2	uc009vje.2	WASH7P	ENST00000438504	653635	n/a	
uc009vjf.2	uc009vjf.2	WASH7P	ENST00000438504	653635	n/a	
uc009vjb.1	uc009vjb.1	WASH7P	ENST00000438504	653635	n/a	
uc001aak.3	uc001aak.3	FAM138F NR_026820	ENST00000417324	641702	n/a	
uc001aal.1	uc001aal.1	OR4F5 NM_001005484	ENST00000335137	79501	n/a	
uc021oeg.2	uc021oeg.2	LOC729737 NR_039983	ENST00000493797	729737	n/a	
uc001aaq.2	uc001aaq.2	DQ597235	n/a	n/a	n/a	
uc001aar.2	uc001aar.2	DQ599768	n/a	n/a	n/a	
uc021oeh.1	uc021oeh.1	LOC100133331	ENST00000440038	100133331	n/a	
uc009vjk.2	uc009vjk.2	LOC100133331	ENST00000440038	100133331	n/a	
uc021oei.1	uc021oei.1	LOC388312	n/a	n/a	n/a	
uc001aaau.3	uc001aaau.3	LOC100132062 NR_028325	ENST00000440038	100132062	n/a	
uc010nxu.2	uc010nxu.2	OR4F29 NM_001005277	ENST00000426406	729759	OR4F3,	
uc001aax.1	uc001aax.1	BC036251	n/a	n/a	n/a	
uc021oej.1	uc021oej.1	JA429830	n/a	n/a	n/a	
uc021oek.1	uc021oek.1	JA429831	n/a	n/a	n/a	
uc021oel.1	uc021oel.1	JB137814	n/a	n/a	n/a	
uc001abb.3	uc001abb.3	M37726	n/a	n/a	n/a	
uc010nxv.2	uc010nxv.2	OR4F29 NM_001005277	ENST00000332831	729759	n/a	
uc001abe.4	uc001abe.4	LOC100133331 NR_028327	n/a	100133331	LOC100133331,	
uc001abi.2	uc001abi.2	DQ575786	n/a	n/a	n/a	
uc001abj.3	uc001abj.3	DQ599872	n/a	n/a	n/a	
uc009vjm.3	uc009vjm.3	LOC100133331	n/a	100133331	n/a	
uc010nxw.2	uc010nxw.2	DQ575786	n/a	n/a	n/a	
uc001abl.3	uc001abl.3	DQ599872	n/a	n/a	n/a	
uc002khh.3	uc002khh.3	LOC100133331	ENST00000416385	100133331	LOC100133331,	
uc001abm.2	uc001abm.2	AK310751	ENST00000416385	n/a	n/a	
uc001abo.3	uc001abo.3	LOC100288069 NR_033908	ENST00000428504	100288069	n/a	
uc010nxx.2	uc010nxx.2	LINC00115 NR_024321	ENST00000473798	79854	LINC00115	
uc031pjj.1	uc031pjj.1	LOC643837 NR_047526	ENST00000415295	643837	n/a	
uc001abp.2	uc001abp.2	LOC643837 NR_047520	ENST00000445118	643837	n/a	
uc021oem.2	uc021oem.2	LOC643837 NR_047522	ENST00000445118	643837	n/a	
uc009vjn.2	uc009vjn.2	LOC643837 NR_047523	ENST00000445118	643837	n/a	

Nota: l'output verrà scaricato in un file con il nome specificato al punto 3, oppure visualizzato a schermo (come nell'esempio mostrato) nel caso in cui nessun nome sia stato indicato.

parte II:

importare e visualizzare il risultato ottenuto dal Table Browser in un foglio di calcolo

Nota: Per questi passaggi il dettaglio dipenderà dal tipo di software (Office EXCEL, libreOffice, etc), versione e sistema operativo (Windows, Mac, linux) utilizzati. Ma i passaggi concettuali da fare sono:

1. apri un nuovo foglio di lavoro
2. dal menù seleziona la opzione: inserisci nuovo foglio da un file
3. caricare il file "gene_ids_map__ucsc.hg19.txt" (output scaricato dal UCSC genome browser) specificando dal menù di importazione il separatore di testo è il tab ("\t")
4. salva il foglio di calcolo appena creato con il nome "gene_ids_map__ucsc.hg19.xls"

	A	B	C	D	E	F	G
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							
23							
24							
25							
26							
27							
28							
29							
30							
31							
32							
33							
34							
35							
36							
37							
38							
39							

Insert Sheet

Position

- ☒ Before current sheet
- ☐ After current sheet

Sheet

- ☐ New sheet
 - No. of sheets: 1
 - Name: Sheet4
- ☒ From file

Text Import - [gene_ids_map_ucsc.hg19.txt]

Import

Character set: Unicode (UTF-8)

Language: Default - English (USA)

From row: 1

Separator options

- ☐ Fixed width
- ☒ Separated by
 - ☒ Tab
 - ☐ Semicolon
 - ☐ Merge delimiters
 - ☐ Comma
 - ☐ Space
 - ☐ Other:

Text delimiter: "

Other options

- ☐ Quoted field as text
- ☐ Detect special numbers

Fields

Column type

	Standard	Standard	Standard
1	#hg19.knownGene.name	hg19.knownToEnsembl.value	hg19.ensGene.name2
2	uc001aaa.3	ENST00000456328	ENSG00000223972
3	uc010nxr.1	ENST00000456328	ENSG00000223972
4	uc010nxq.1	ENST00000518655	ENSG00000223972
5	uc009vis.3	ENST00000423562	ENSG00000227232
6	uc009vjc.1	ENST00000541675	ENSG00000227232
7	uc009vjd.2	ENST00000488147	ENSG00000227232