Esercitazione UCSC – Table Browser.

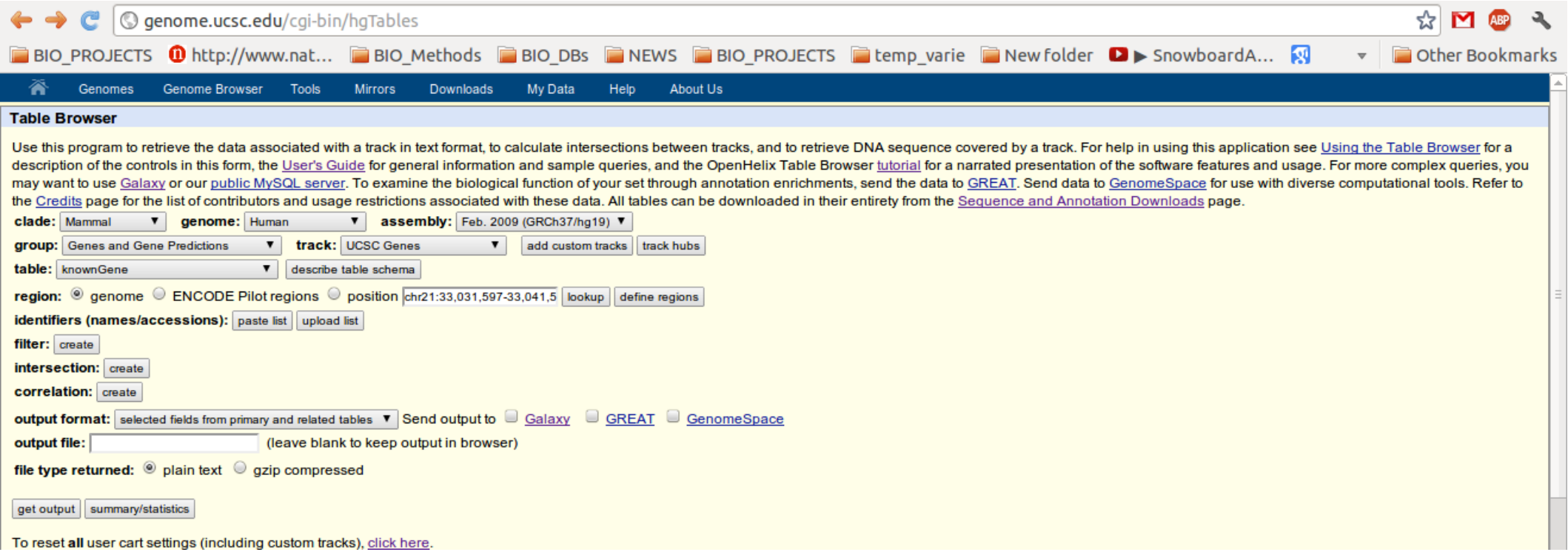
Obiettivi dell'esercitazione:

(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).

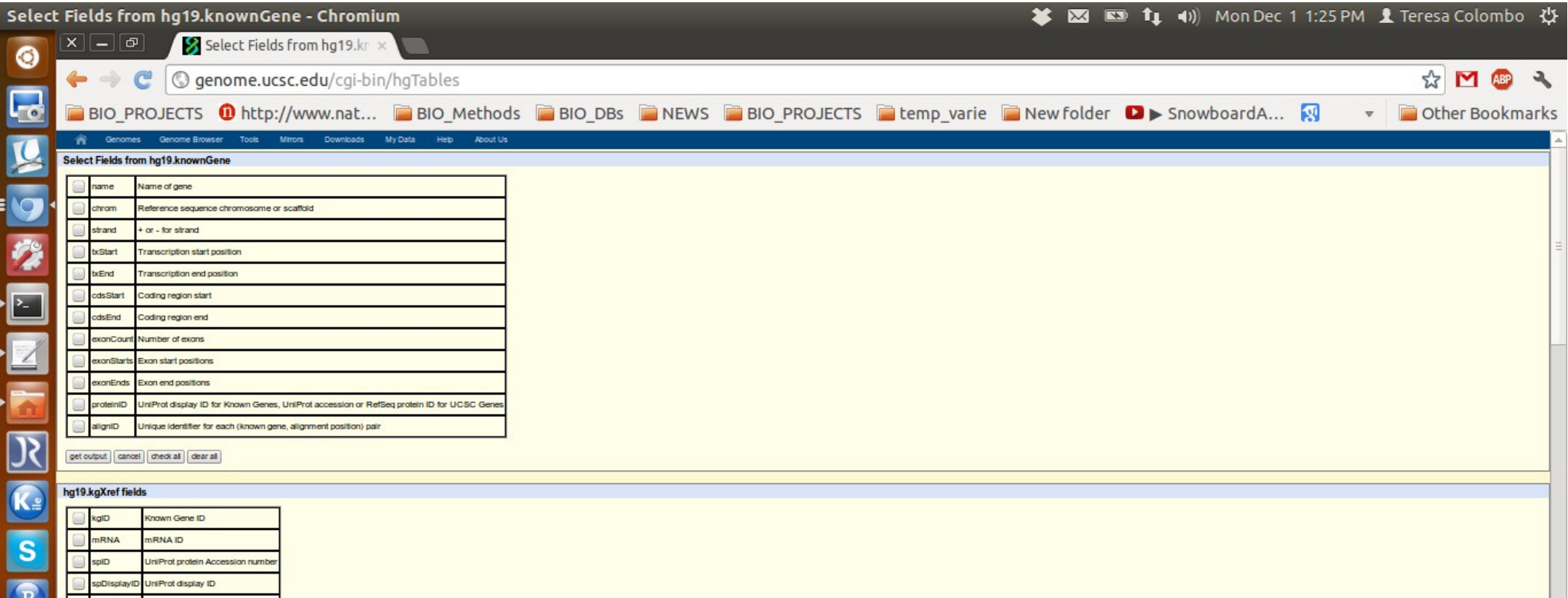
1. apri una pagina del UCSC genome browser http://genome.ucsc.edu
2. dal menù principale --> Tools --> Table Browser
3. definisci i criteri di ricerca come segue:
   clade:Mammal genome:Human assembly: Feb.2009(GRCh37/hg19)
   group:Genes and Gene Predictions track: UCSC Genes
   table:knownGene
   region:genome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



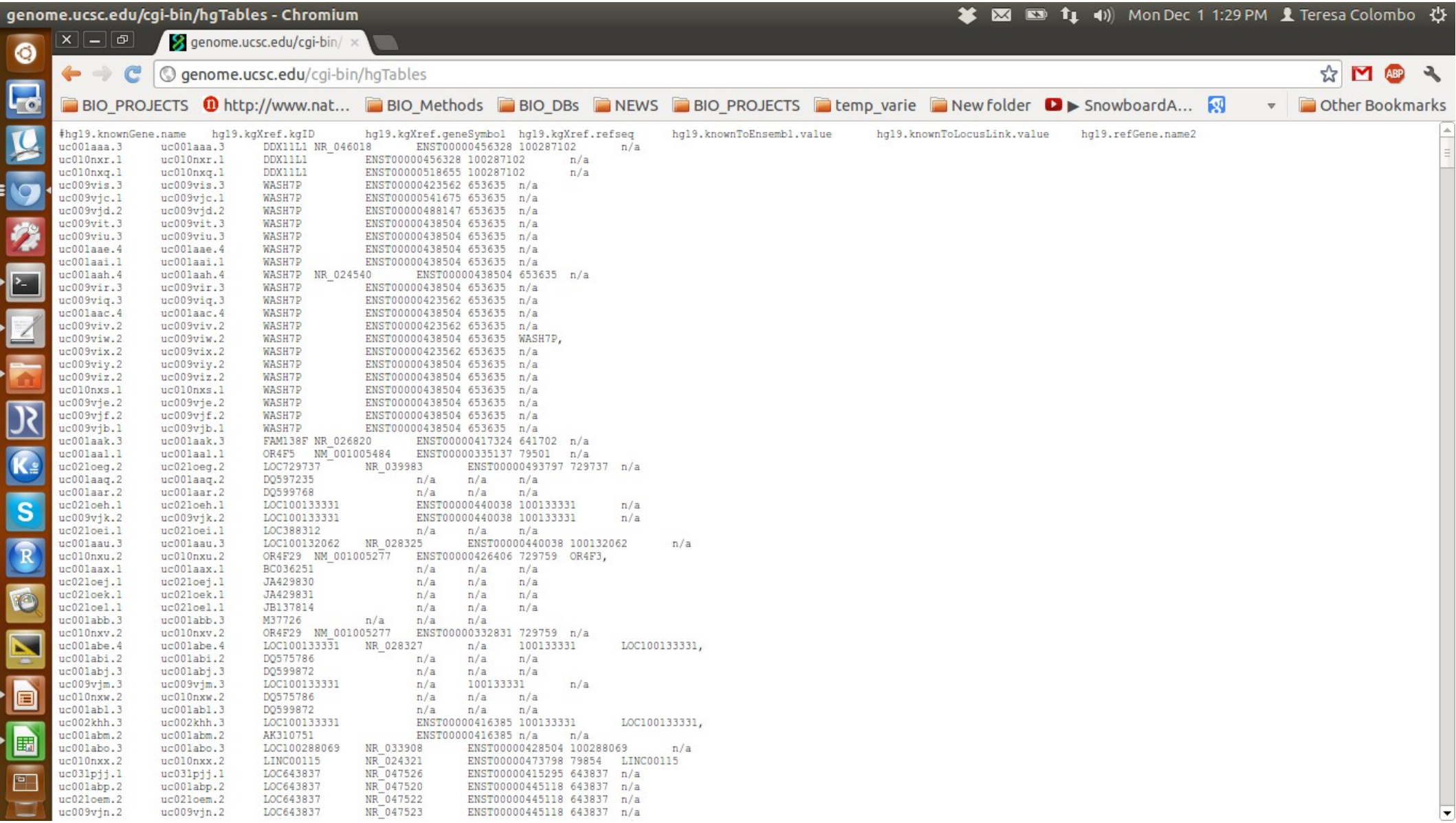
4. premi "get output" --> verrai indirizzato ad una nuova pagina con la possibilità di selezionare vari campi -sia dalla tabella "knownGene" che da altre tabelle correlate - da riportare nel file di output;

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;



4.c premi il pulsante "get output".



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.

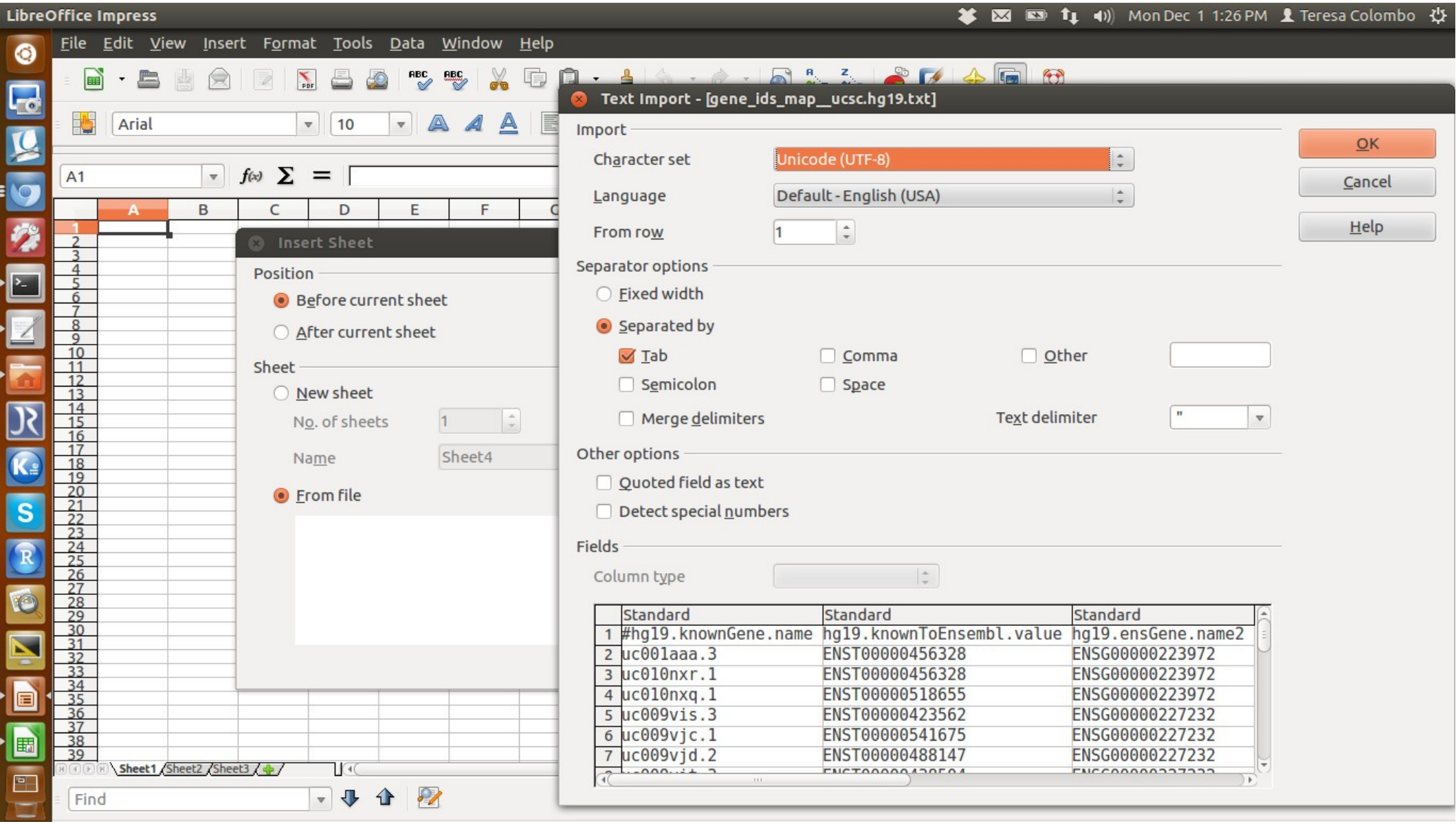
parte II:
importare e visualizzare il risultato ottenuto dal Table Browser in un foglio di calcolo

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"


LibreOffice Impress
Mon Dec 1 1:26 PM
Teresa Colombo
File Edit View Insert Format Tools Data Window Help
Arial
10
A1
Insert Sheet
Position
Before current sheet
After current sheet
Sheet
New sheet
No. of sheets
Name
Sheet4
From file
Sheet1 Sheet2 Sheet3
Find
Text Import - [gene_ids_map__ucsc.hg19.txt]
Import
Character set
Unicode (UTF-8)
Language
Default - English (USA)
From row
1
OK
Cancel
Help
Separator options
Fixed width
Separated by
Tab
Comma
Other
Semicolon
Space
Merge delimiters
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