



NimbleGen Sequence Capture Custom Designs

Guide to Submitting Your Target Sequence

Outline

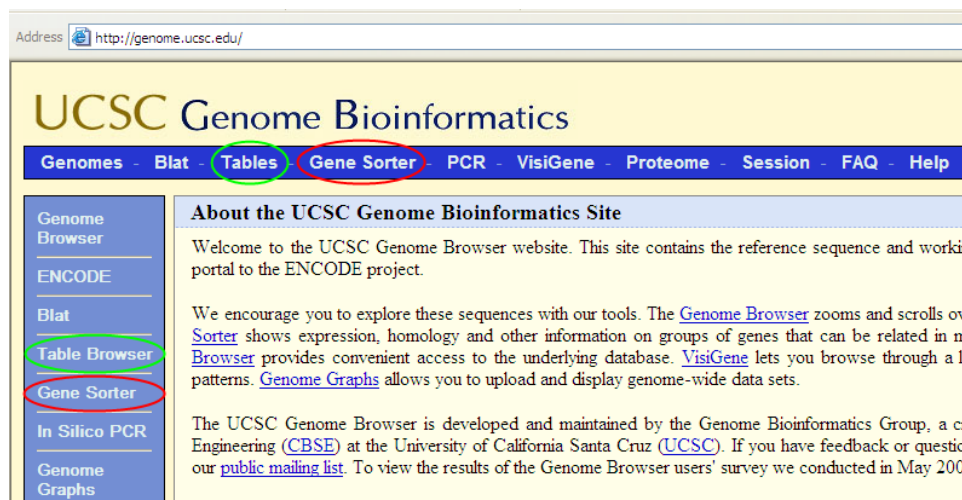
Generating a list of genomic coordinates for a NimbleGen Sequence Capture array is an important part of setting up your custom array design. Roche NimbleGen strongly suggests that you use the UCSC Genome Browser to successfully generate a list of genome coordinates for use in specifying your capture regions. The Tutorial in this guide is designed to assist you in generating both a gene list and this coordinate list, and the Quick Instructions on page 11 offer steps for generating a list of exons.

Tutorial

Step 1. Generate a Gene List

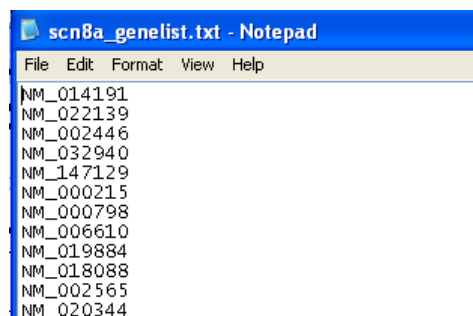
- 1 If you already have a gene list, skip to “Step 2. Convert Gene List to Chromosomal Coordinates,” beginning on page 5.
- 2 If you do not have a gene list, use your browser to go to the home page for the UCSC Genome Browser: <http://genome.ucsc.edu>.

From this home page, you will use the options for **Table Browser**, accessible from either the Tables link in the main menu bar at the top of the page or the Table Browser link in the link list at left (circled in **green**), and **Gene Sorter**, accessible from either the main menu bar or the links list (circled in **red**).

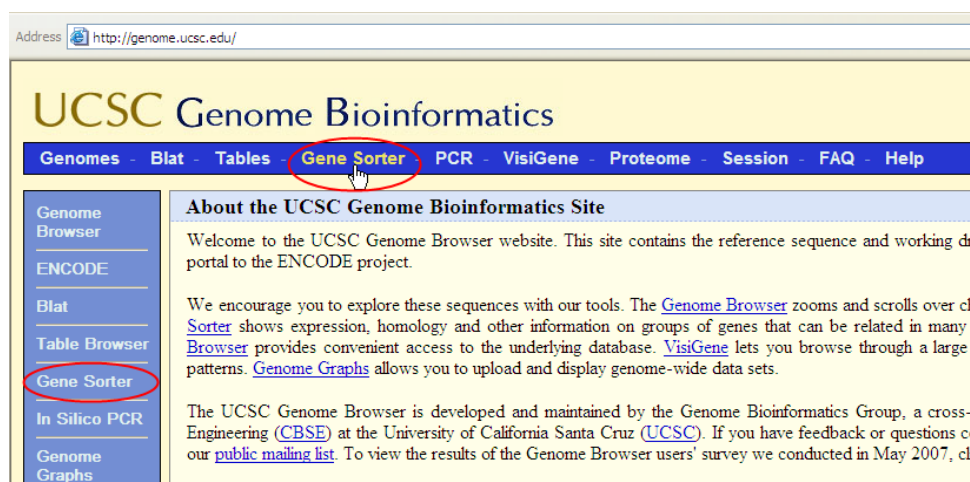


The **Table Browser** is useful to the Sequence Capture customer because it can be used to generate a .bed file of chromosome locations and exon start and stop coordinates, which are the elements needed by the array design team to create a custom Sequence Capture design. It also can be used to display a page of these coordinates, which can be copied and pasted into other applications, and manipulated there. The **Gene Sorter** utility allows you to enter a gene name and search for genes that have similarities to a search term and can be used to generate a gene list with one gene per line.

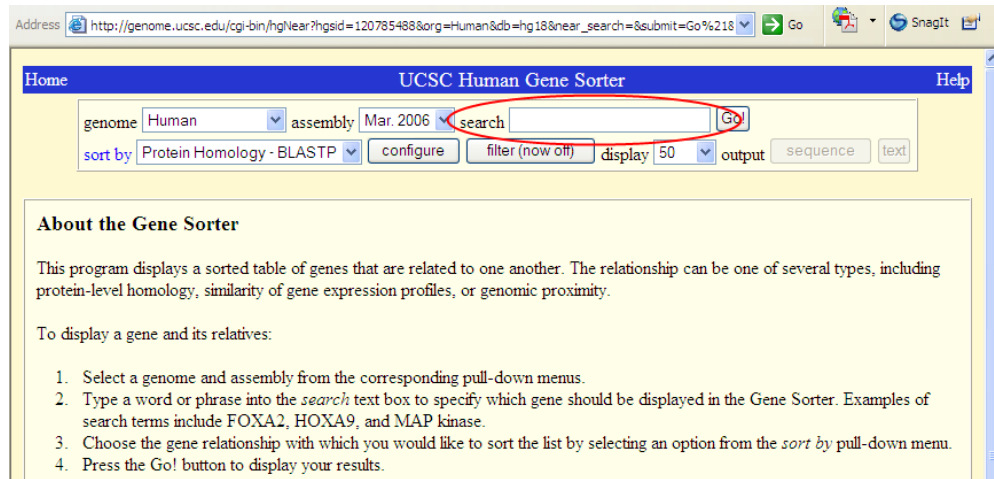
Your list of genes will ideally be comprised of only of accession (NM) numbers, but may also include gene names. The gene list should be in a tab-delimited, text file format, with one gene per line. Viewed in Notepad, this kind of file looks like the following:



- 3 To access the Gene Sorter tool, click a Gene Sorter link on the UCSC home page.

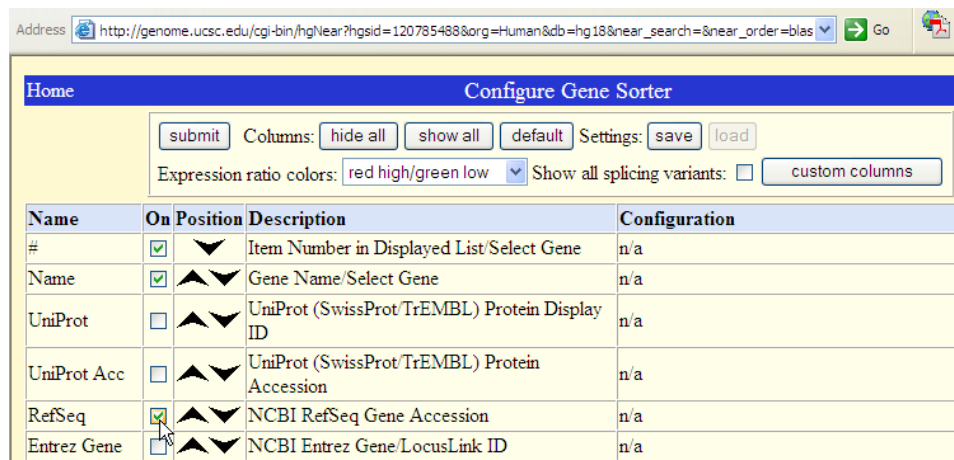


Initially an About the Gene Sorter page appears with the Gene Sorter tool at the top of the page and instructional guidelines at the bottom. Use this tool to enter parameters to search for similar genes from the curated UCSC database by entering any kind of gene name, accession number, or other term in the *search* field.



Search results will appear in a table with a set of columns that can be customized by clicking **configure**.

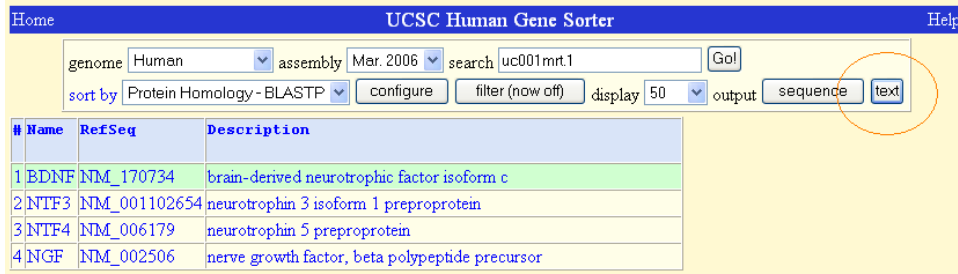
- 4 Click **configure** and select “On” for the RefSeq track. In the example below, the selected tracks include #, Name, RefSeq, and Description (not shown), then click **submit** to return to the main UCSC Human Gene Sorter tool page.



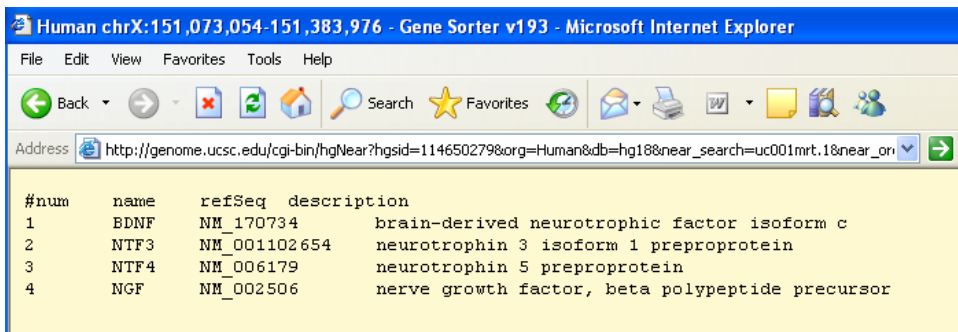
- 5 Enter “BDNF” in the *search* field and select “Protein Homology – BLASTP” in the *sort by* field to set up a search by protein sequence homology.



- 6 Click **Go!** to run the search. The tabular list that results includes columns for each of the selected configuration parameters.

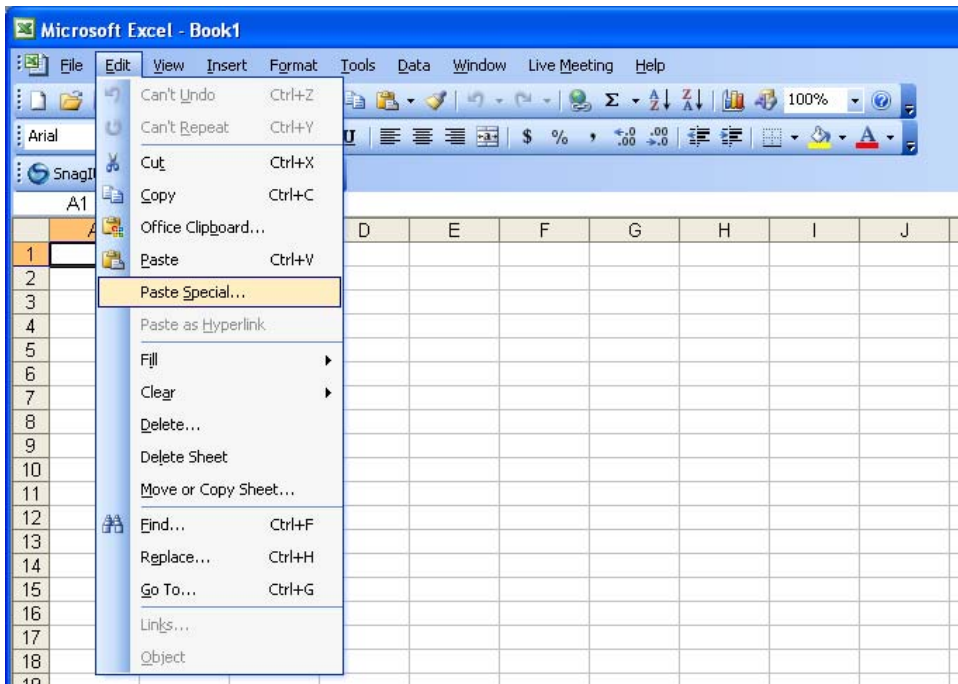


- Click text (circled in orange) to view the list in a tab-delimited text format.

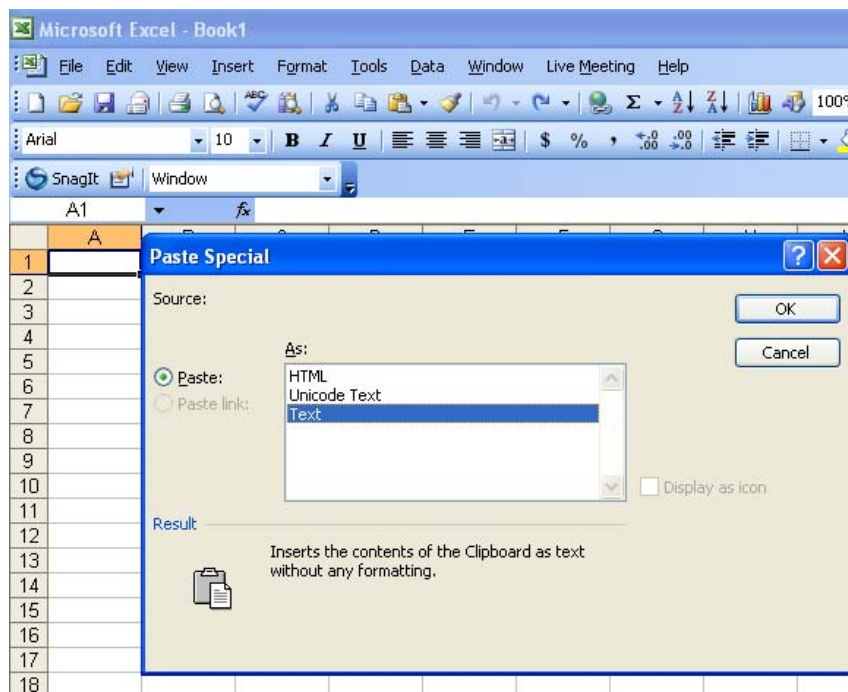


This gene list is now in a format where you can select, copy, and paste into other applications.

- Select and copy all of the text shown on the screen.
- Select Edit -> Paste Special... to paste the output as text into a blank Excel worksheet.



- In the Paste Special dialog, select *Paste As*: "Text" and click OK.



The data will be pasted into the worksheet in a one-value-per-cell format.

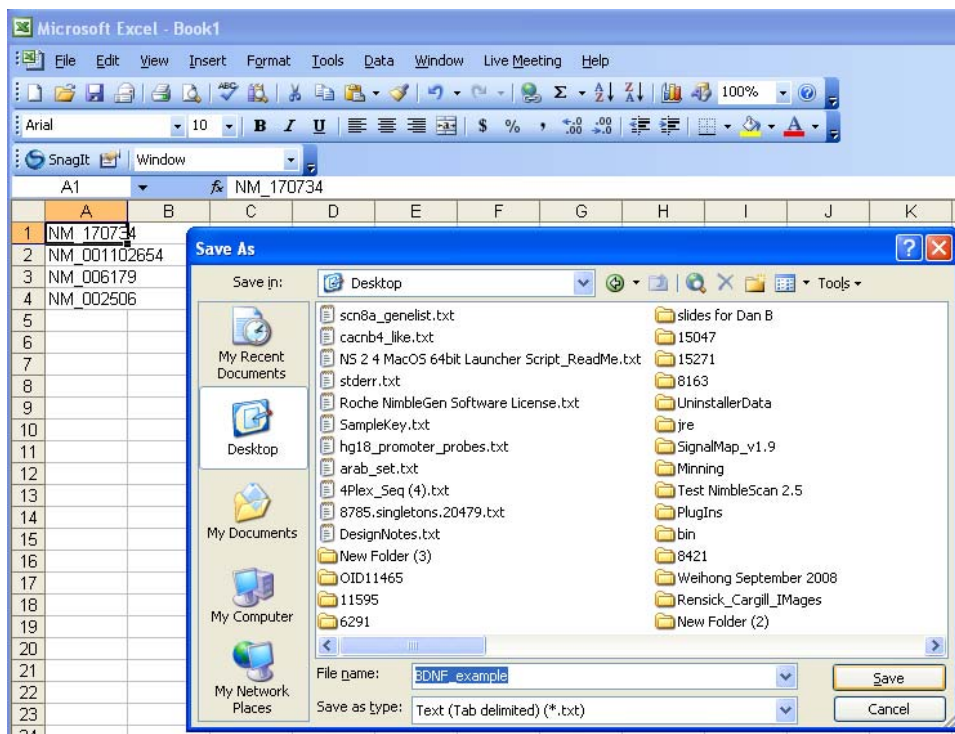
	A	B	C	D	E	F	G	H	I	J
1	#num	name	refSeq	description						
2	1	BDNF	NM_17073	brain-derived neurotrophic factor isoform c						
3	2	NTF3	NM_00110	neurotrophin 3 isoform 1 preproprotein						
4	3	NTF4	NM_00617	neurotrophin 5 preproprotein						
5	4	NGF	NM_00250	nerve growth factor, beta polypeptide precursor						
6										

- 11 Manipulate data as needed using the tools available in Excel, and save the file.

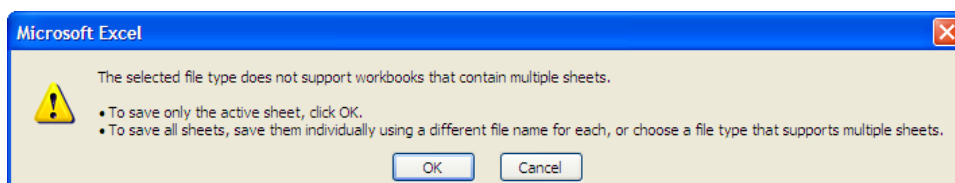
Step 2. Convert Gene List to Chromosomal Coordinates

From the Gene List created in Step 1.11, you will be able to use the accession numbers in the refSeq column to create a list of chromosomal coordinates for each listed gene. This file will ultimately contain the accession number, chromosome number, and exon start and stop locations.

- 1 Starting with the same Excel file from Step 1.11, select and delete all columns *except* refSeq.
- 2 Select File -> Save As... to save the list of accession numbers in a tab-delimited text format (*.txt).

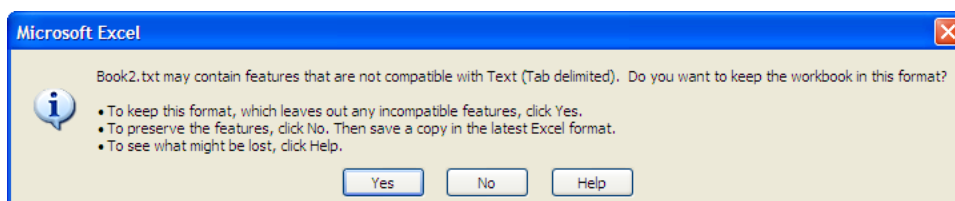


If your Excel file contains more than one worksheet (the default workbook contains three worksheets), you will see a warning dialog that the tab-delimited text file type does not support workbooks that contain multiple sheets.



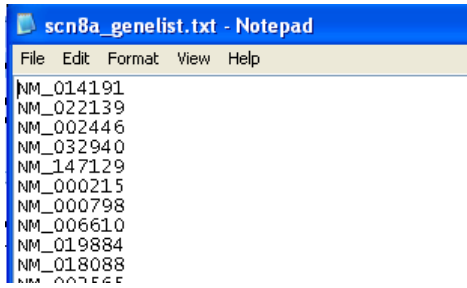
Since you only need to save the active worksheet, click OK.

You will see an information dialog that lets you know that you might lose features that are not compatible with the tab-delimited text format box.

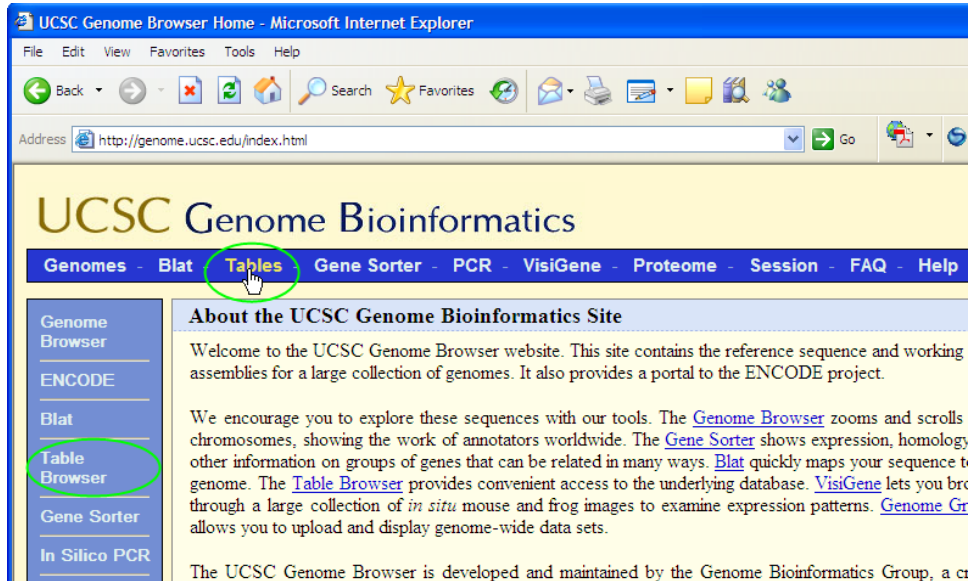


Click Yes. This ensures that the file contains only appropriate text characters.

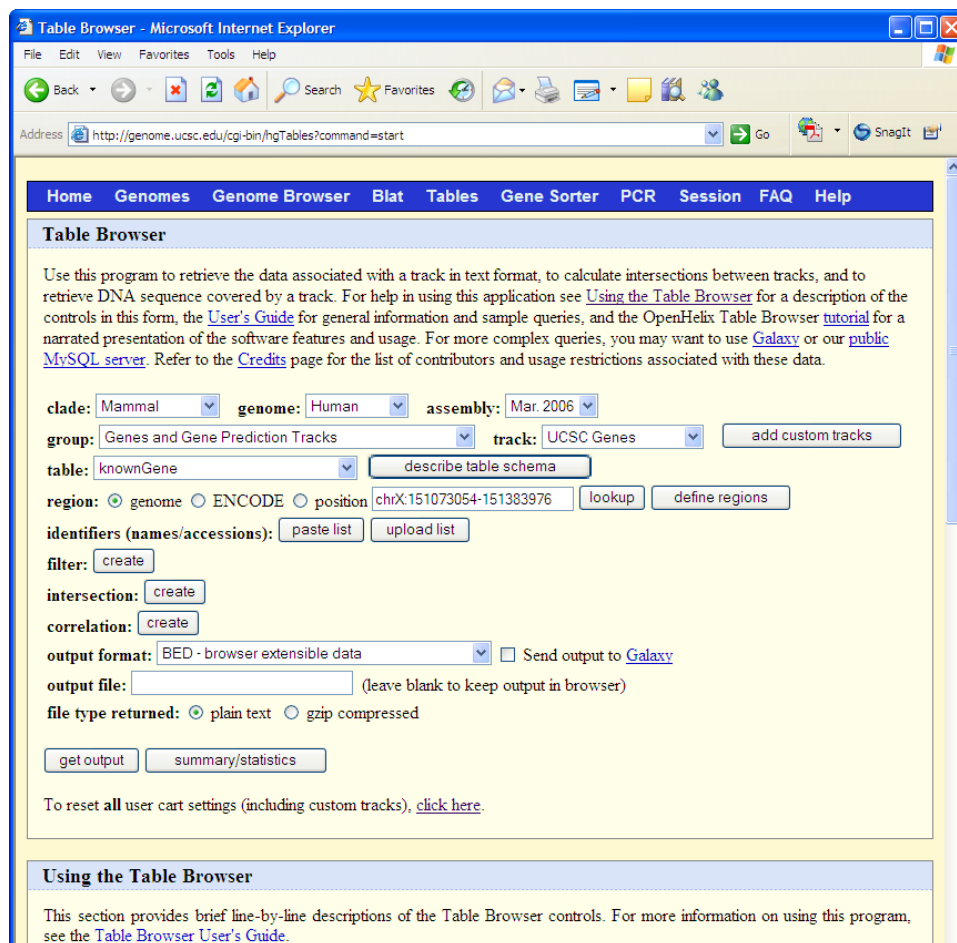
You now have a tab-delimited text file of accession numbers from which you can generate exon coordinates. Viewed in Notepad, this kind of file looks like the following:



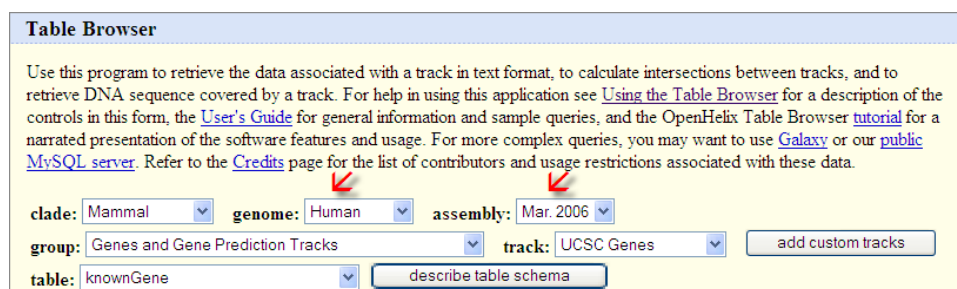
- 3 Select Tables from the main menu bar or Table Browser from the list menu on the front page of the UCSC Genome Browser.



The Table Browser page appears with a description of the tool at the top of the page, the Table Browser tool in the middle, and guidelines for use at the bottom.



- Select the *genome* as “Human” and *assembly* as “Mar. 2006”.



Important: For Roche NimbleGen Sequence Capture service, only Human DNA is accepted for submission. We require the use of the March 2006 assembly of the human genome to generate chromosomal and exon coordinates. Make sure you select **HUMAN** as the genome and **Mar. 2006** as the assembly in the UCSC Table Browser.

Note: We recommend that you reset all user cart settings by clicking the [click here](#) link in the line of text below the **get output** button. The UCSC browser was designed to remember user settings and clicking this link will ensure that you will be using the correct settings rather than incorrect settings that might have been retained from a previous browsing session. Clicking this text link also automatically resets the genome as **HUMAN** and the assembly as **Mar. 2006**.

file type returned: plain text gzip compressed

To reset **all** user cart settings (including custom tracks), [click here](#).

- 5 Select the **output format** as “BED - browser extensible data”.
- 6 The **identifiers (names/accessions)** field offers two clickable buttons: **paste list** and **upload list**. Click **upload list** to transfer the tab-delimited text file containing your gene list saved in Step 2.2. The Upload Identifiers for UCSC Genes page appears.

Upload Identifiers for UCSC Genes

Please enter the name of a file from your computer that contains a space, tab, or line separated list of the items you want to include. The items must be values of the **name** field of the currently selected table, **knownGene**, or the **alias** field of the alias table **kgAlias**. (The “describe table schema” button shows more information about the table fields.) Some example values:

```
uc004egm.2
uc001ojk.1
uc003zri.2
uc001wzn.1
NP_150634
NM_001099402
```

Click **Browse** to locate the gene list text file and then click **submit**. The main Table Browser page appears.

- 7 Click **get output**. You do not need to change any of the rest of the parameters. The Output knownGene as BED page appears.

There are a set of eight *Create one BED record per* data format options.

- 8 Select “Exons plus 0 bases at each end”. The default setting is “Whole Gene”.
- 9 Click get BED.

The output looks like the following:

```

http://genome.ucsc.edu/cgi-bin/hgTables?hgid=114651446&boolshad.hgta_printCustomTrackHeaders=1 - Microsoft I
File Edit View Favorites Tools Help
Back Search Favorites
Address http://genome.ucsc.edu/cgi-bin/hgTables?hgid=114651446&boolshad.hgta_printCustomTrackHeaders=1&hgta_ctNarr
chr1 115630059 115630951 uc001efu.1_exon_0_0_chr1_115630060_r 0 -
chr1 115637770 115637894 uc001efu.1_exon_1_0_chr1_115637771_r 0 -
chr1 115682347 115682380 uc001efu.1_exon_2_0_chr1_115682348_r 0 -
chr1 115630213 115630939 uc009wgx.1_exon_0_0_chr1_115630214_r 0 -
chr11 27633017 27636708 uc001mrt.1_exon_0_0_chr11_27633018_r 0 -
chr11 27677504 27677756 uc001mrt.1_exon_1_0_chr11_27677505_r 0 -
chr12 5411540 5411769 uc001qnk.2_exon_0_0_chr12_5411541_f 0 +
chr12 5473620 5474726 uc001qnk.2_exon_1_0_chr12_5473621_f 0 +
chr19 54256208 54257078 uc002pmf.2_exon_0_0_chr19_54256209_r 0 -
chr19 54258807 54258936 uc002pmf.2_exon_1_0_chr19_54258808_r 0 -
chr19 54256433 54257066 uc010emr.1_exon_0_0_chr19_54256434_r 0 -

```

- 10 Select and copy all of the text shown on the screen.
- 11 Select Edit -> Paste Special... to paste the output as text into a blank Excel worksheet.
- 12 In the Paste Special dialog, select *Paste As*: “Text” and click OK. The data will be pasted into the worksheet in a one-value-per-cell format.

	A	B	C	D	E	F	G
1	chr1	115630059	115630951	uc001efu.1_exon_0_0_chr1_115630060_r	0	-	
2	chr1	115637770	115637894	uc001efu.1_exon_1_0_chr1_115637771_r	0	-	
3	chr1	115682347	115682380	uc001efu.1_exon_2_0_chr1_115682348_r	0	-	
4	chr1	115630213	115630939	uc009wgx.1_exon_0_0_chr1_115630214_r	0	-	
5	chr11	27633017	27636708	uc001mrt.1_exon_0_0_chr11_27633018_r	0	-	
6	chr11	27677504	27677756	uc001mrt.1_exon_1_0_chr11_27677505_r	0	-	
7	chr12	5411540	5411769	uc001qnk.2_exon_0_0_chr12_5411541_f	0	+	
8	chr12	5473620	5474726	uc001qnk.2_exon_1_0_chr12_5473621_f	0	+	
9	chr19	54256208	54257078	uc002pmf.2_exon_0_0_chr19_54256209_r	0	-	
10	chr19	54258807	54258936	uc002pmf.2_exon_1_0_chr19_54258808_r	0	-	
11	chr19	54256433	54257066	uc010emr.1_exon_0_0_chr19_54256434_r	0	-	
12							

- 13 Select and delete unwanted columns *except* Chromosome, Exon Start, and Exon Stop coordinates.
- 14 Select File -> Save As... to save the coordinates list in a tab-delimited text format (*.txt).



If your Excel file contains more than one worksheet (the default workbook contains three worksheets), you will see a warning dialog that the tab-delimited text file type does not support workbooks that contain multiple sheets. Since you only need to save the active worksheet, click OK.

You will see an information dialog that lets you know that you might lose features that are not compatible with the tab-delimited text format box. Click Yes. This ensures that the file contains only appropriate text characters.

- 15 You now have a tab-delimited text file that is ready to submit to the Roche NimbleGen array design team for Sequence Capture.

Quick Instructions for

Generating a .bed file of exons from the UCSC Genome Browser

The quick instructions for generating a list of exons from the UCSC browser are as follows:

- 1 Open <http://genome.ucsc.edu> and click **Tables** in the blue menu bar at page top (or **Table Browser** from the link list).
- 2 Select the *genome* as “Human” and *assembly* as “Mar. 2006”.
- 3 Add identifiers by using either **paste list** or **upload list** buttons, then click **submit** to add identifiers and return to the Table Browser page.
- 4 Click **get output** to open the Output knownGene as BED page and select the appropriate line under *Create one BED record per:* as “Whole Gene” or “Exons plus X bases at each end”.
- 5 Click **get BED** to view the results in text format within your browser.
- 6 Copy and paste all results as text into an Excel worksheet using **Edit -> Paste Special** and save the file.

Technical Support

We hope this information is helpful. If you have any questions about this process, please contact your local technical support team. In North American, email Biochemts.us@roche.com.

If you have questions about Sequence Capture arrays, contact your Roche NimbleGen Account Manager or Roche Microarray Technical Support. Go to www.nimblegen.com/arrayssupport for contact information.



HIGH - DEFINITION GENOMICS

© 2009 Roche NimbleGen, Inc. All Rights Reserved.
03/09



Roche NimbleGen, Inc.
504 S. Rosa Road
Madison, WI 53719 USA

For life science research only.
NIMBLEGEN is a trademark of Roche. All other trademarks are the property of their respective holders.