# TargetSeq™ Custom Enrichment Kit

Guide for TargetSeq™ Custom Enrichment Kit Orders

**Publication Part Number** 4471870 Rev. A **Revision Date** August 2011

This quick reference guide covers:



## Ordering TargetSeq™ Custom Enrichment Kits

TargetSeq™ Custom Enrichment Kits provide a convenient method to enrich for targeted genomic regions. This guide helps you generate the data required to place your order and review the resulting design provided by Life Technologies.

### About the ordering process

TargetSeq™ Custom Enrichment Kit ordering and production is a multi-phase process. The workflow (see right) describes the entire process from beginning to end. The design and approval phases are performed via email, and they require customer participation.

**Note:** After a panel design is approved, production begins and the order is irrevocable.

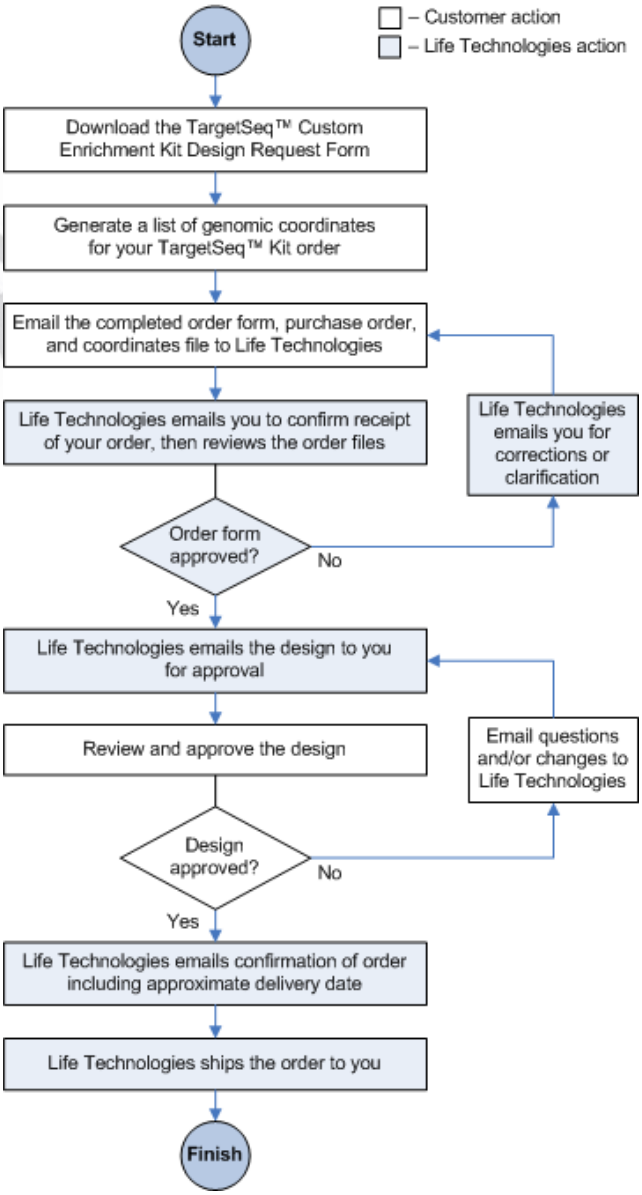
### About the order and design files

You will exchange design files with Life Technologies at least twice:

- First, when submitting your coordinates file, and
- Again, when Life Technologies submits the panel design for your approval.

The following table describes the files that you will send to and receive from Life Technologies during the design phase.

| File | Description |
|---|---|
| Coordinates file (.txt) | A customer-created, tab-delimited text file that contains the design specifications for a custom design order. |
| Browser Extensible Data file (.bed) | Design files created by Life Technologies that can be read using the UCSC Genome Browser. |
| Summary Design file (.txt) | A text file that describes the properties of your TargetSeq™ Custom Design. |






life technologies™

## Submitting custom designs to Life Technologies

This section describes how to use the UCSC Genome Browser to generate a list of genome coordinates for use in specifying your capture or tiling regions. You will email the resulting coordinates file along with your purchase order and the completed TargetSeq™ Custom Enrichment Kit Library Design Request Form to Life Technologies to begin the design phase.

**Note:** This section uses HG18 as an example genome. Different builds available in the tool may produce slight differences in output.

### About the UCSC Genome Browser

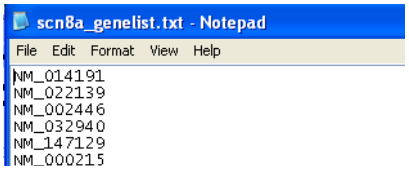
You will use the following features of the UCSC Genome Browser website in this procedure.

| Feature | Description |
|---|---|
| Table Browser | 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 design. The feature can also display a page of these coordinates, which can be copied and pasted into other applications. |
| Gene Sorter | Used to search for a gene name or for genes that have similarities to a search term and can be used to generate a gene list with one gene per line. |

### Step 1: Generate a Gene List

In this procedure you will use the UCSC Genome Browser to generate a gene list. The gene list (see right) is a tab-delimited, text file that contains a list of accession (NM) numbers and gene names (optional) for your regions of interest. Each line of the file lists an accession number and an optional gene name for a single gene of interest, separated by a tab character.

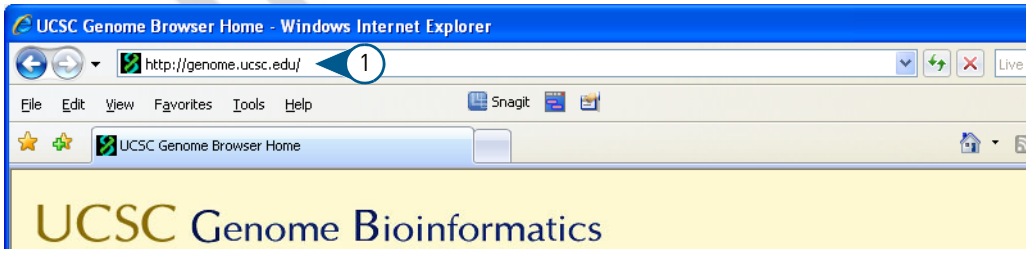
**Note:** If necessary, you can create the gene list file manually.




**1** **Open the UCSC Genome Browser**

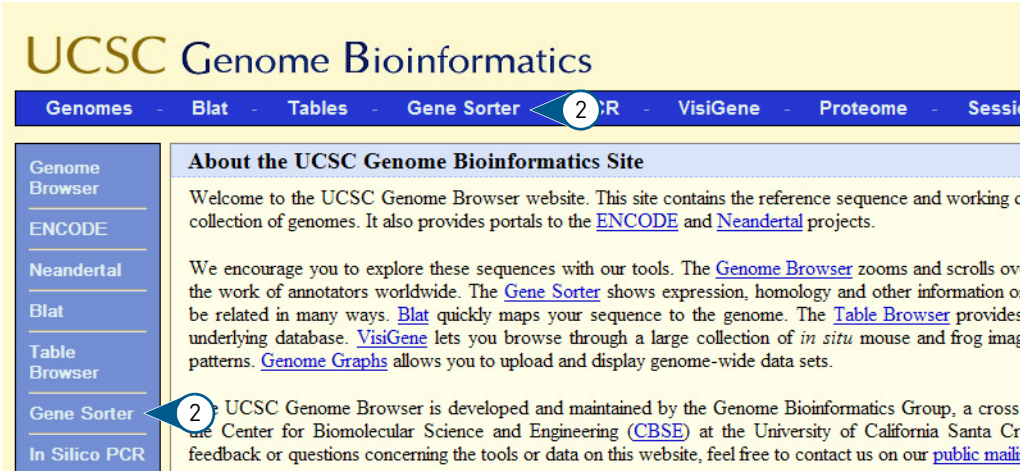
Open the UCSC Genome Browser website (**http://genome.ucsc.edu**).

If you already created a gene list, go to "Step 2: Convert Gene List to Chromosomal Coordinates" on page 6.



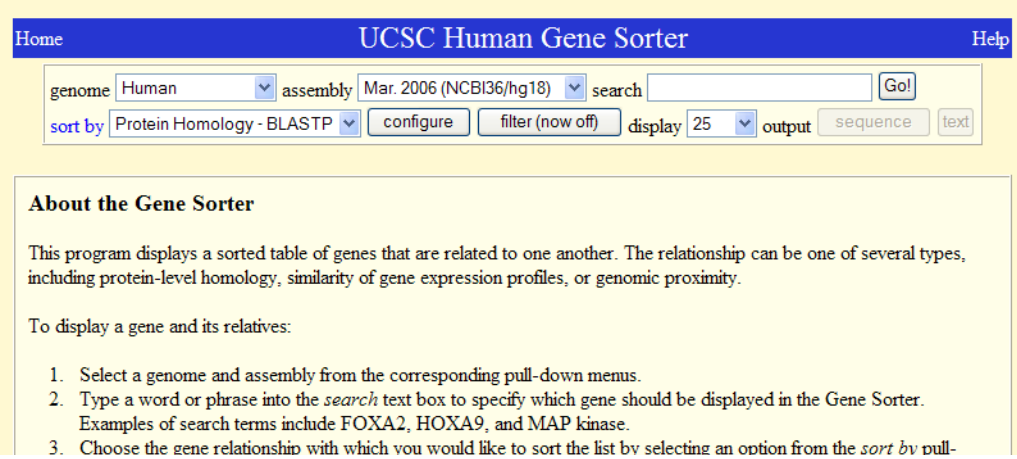

**2** **Open the Gene Sorter**

In the UCSC Genome Browser home page, click **Gene Sorter**.



## 2 Open the Gene Sorter *(continued)*

Initially the "About the Gene Sorter" page appears; this page displays the Gene Sorter tool at the top of the page and the instructions at the bottom. Use the Gene Sorter tool 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. The Gene Sorter page displays search results in a table with a set of columns that can be customized by clicking configure.



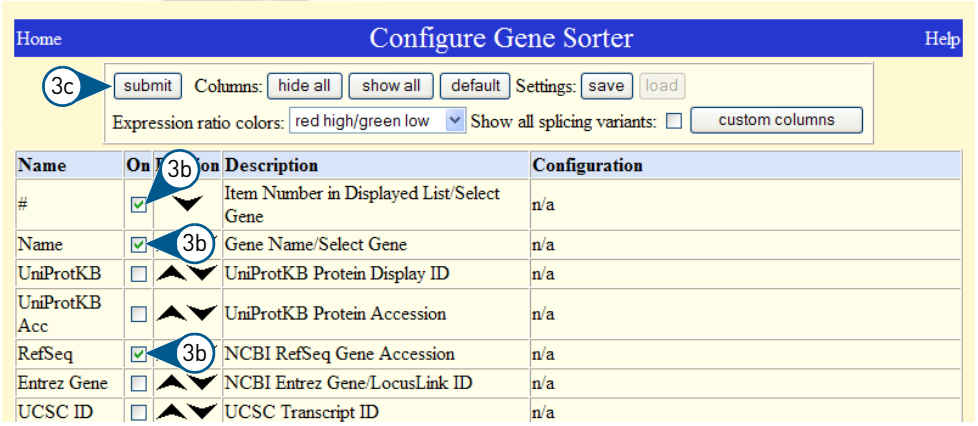

## 3 Configure the search options

a. In the UCSC Human Gene Sorter page, click **configure**.

b. In the Configure Gene Sorter page, select the check box in the **On** column for the RefSeq row.

**Note:** In the example below, selected tracks include #, Name, RefSeq, and Description (not shown).

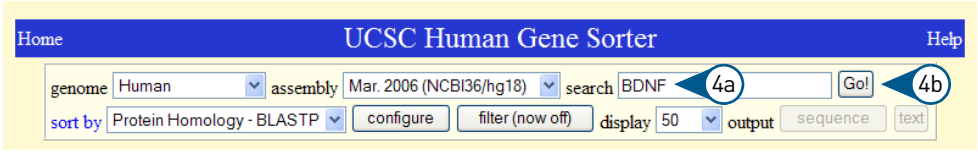
c. Click **submit** to return to the main UCSC Human Gene Sorter page.




## 4 Perform the search

a. Enter one or more keywords in the search field, then configure the search parameters appropriately.

For example, enter "BDNF" in the search field, then select "Protein Homology - BLASTP" in the sort by field to set up a search by protein sequence homology.

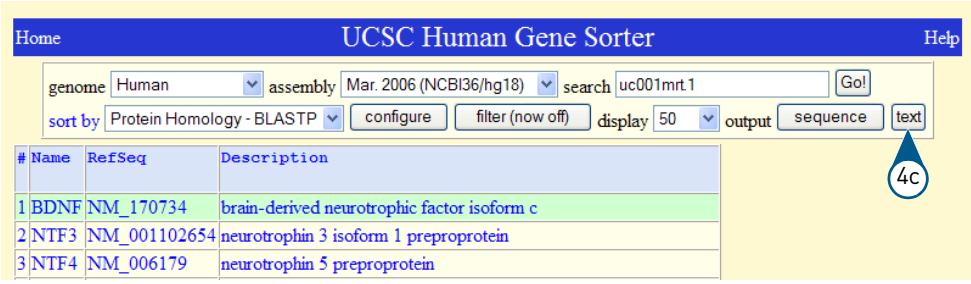



b. Click **Go** to run the search.

The tool returns the search results in a tabular list that includes columns for each of the selected configuration parameters.

**4** Perform the search *(continued)*

**c.** Click **text** to view the search results in a tab-delimited text format.




The gene list is now in a format that you can copy and paste into other applications.

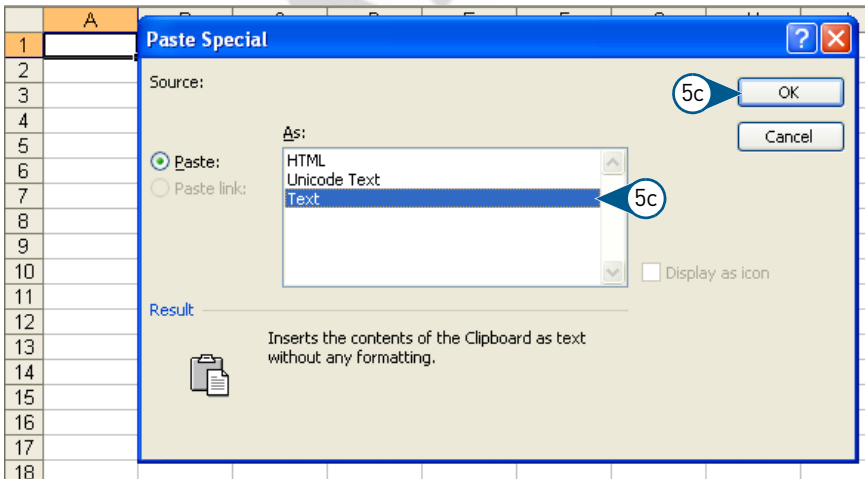
```
#num    name    refSeq  description
1       BDNF    NM_170734       brain-derived neurotrophic factor isoform c
2       NTF3    NM_001102654    neurotrophin 3 isoform 1 preproprotein
3       NTF4    NM_006179       neurotrophin 5 preproprotein
4       NGF     NM_002506       nerve growth factor, beta polypeptide precursor
```

**5** **Copy and paste the search results into Microsoft Excel**

**a.** Select and copy all of the text shown on the screen.

**b.** In Microsoft Excel, select **Edit ▸ Paste Special** to paste the output into a blank worksheet.

**c.** In the Paste Special dialog, select **Text** in the Paste As field, then click **OK** to paste the data into the worksheet in a one-value-per-cell format.




**6** **Edit and save the data**

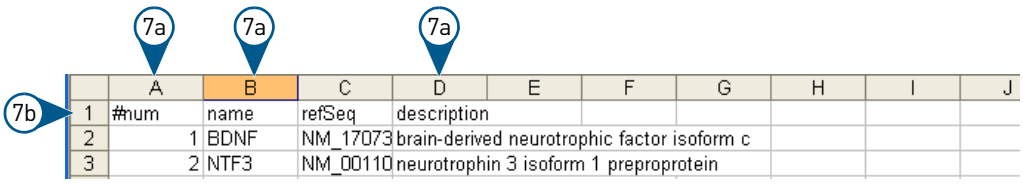
Edit data as needed using Excel, then save the file (select **File ▸ Save**).

**7** **Remove all but the refSeq column**

**a.** Remove all columns, except for refSeq, from the spreadsheet.

For example, remove columns A, B, and D so that only the refSeq column remains.

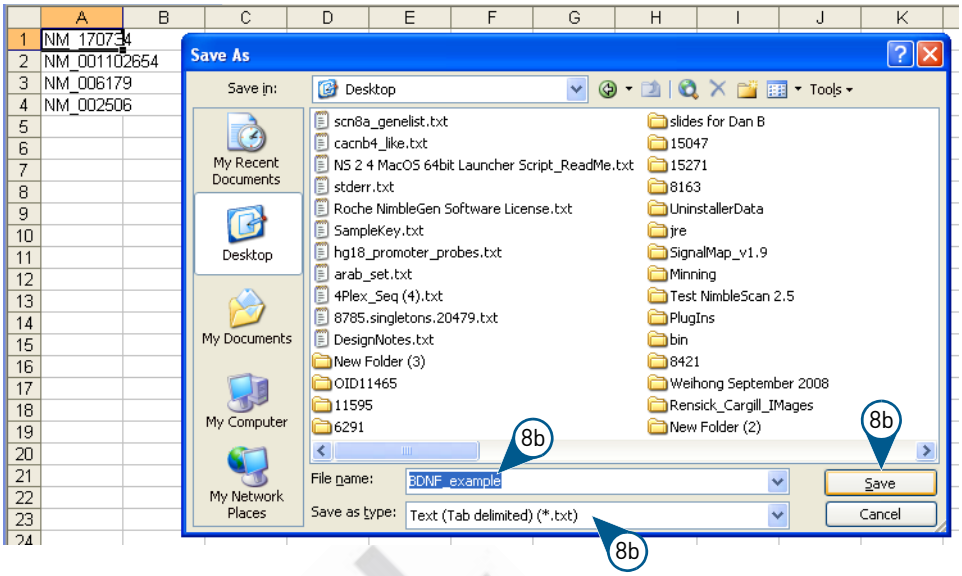
**b.** Remove the heading row (row 1) from the spreadsheet.



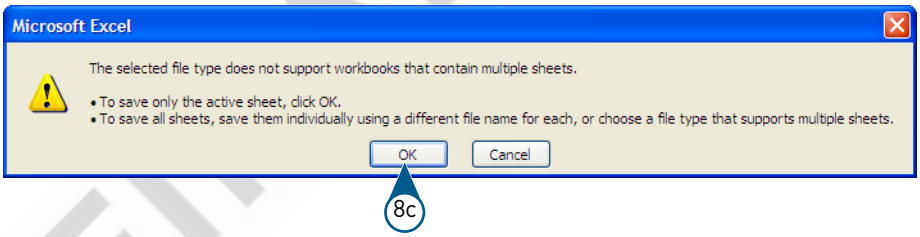
## 8 Save the gene list file

a. Select **File ▸ Save As** to save the list of accession numbers in a tab-delimited text format (*.txt).

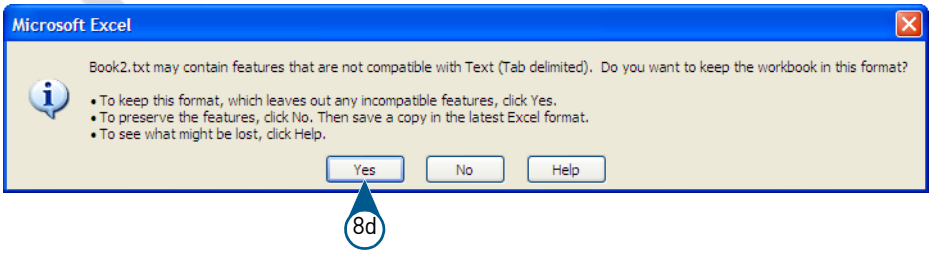
b. In the Save As dialog box, enter a name for the file, select **Text (Tab-delimited)(*.txt)** from the Save as type drop-down list, then click **Save**.




c. If Microsoft Excel displays the following warning, click **OK**.

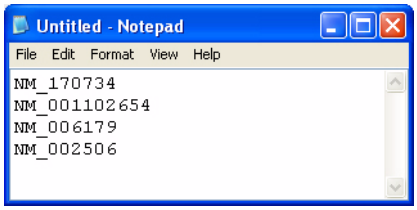



d. If Microsoft Excel displays the following warning, click **Yes**.




You now have a tab-delimited gene list file that contains accession numbers that you can use to generate exon coordinates for your order.

When viewed in Microsoft Notepad, the file appears as shown below.

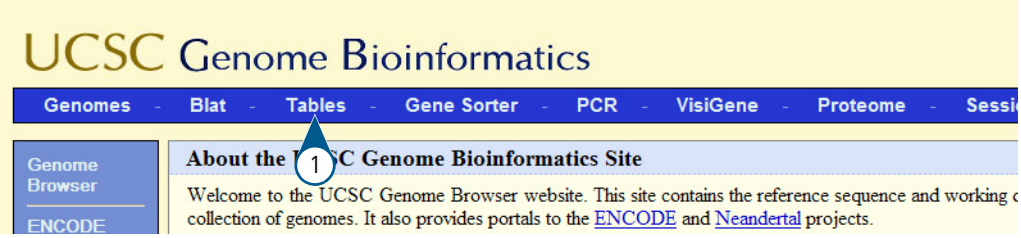


## Step 2: Convert Gene List to Chromosomal Coordinates

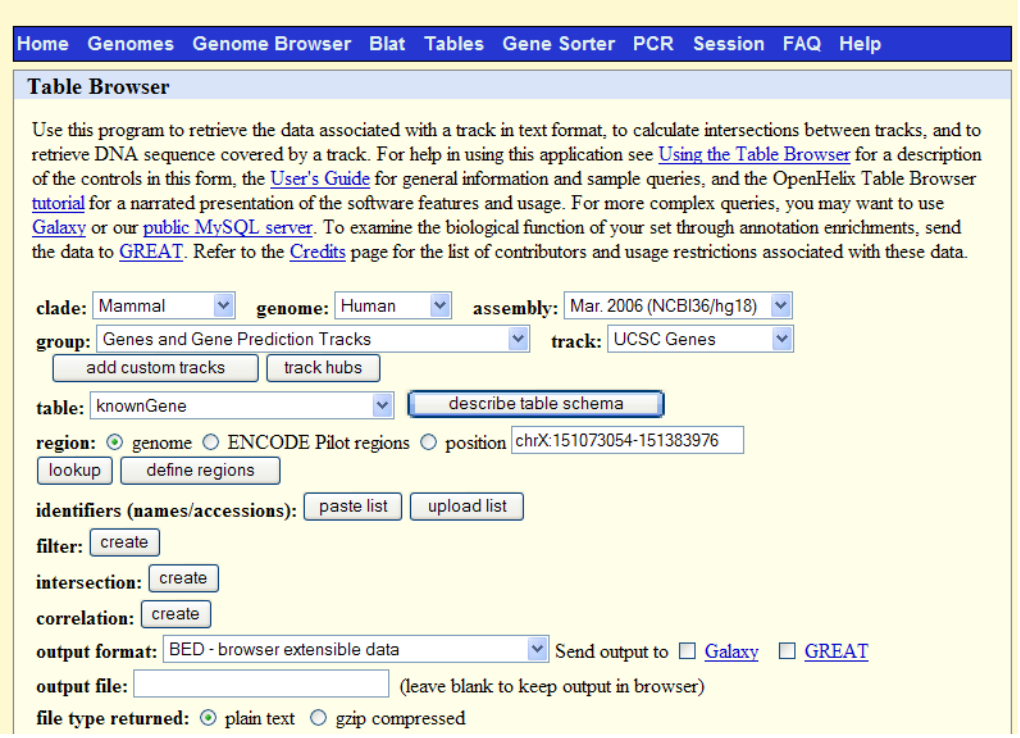
In this procedure, you will use the accession numbers in the Gene List that you created in the previous section to generate a list of chromosomal coordinates for the listed genes. The resulting file, which you will submit with your order, will contain the accession number, chromosome number, and exon start and stop locations for your order.

**1** **Open the Table Browser tool**

In the UCSC Genome Browser, select **Tables** from the main menu bar (or **Table Browser** from the list menu on the home page).

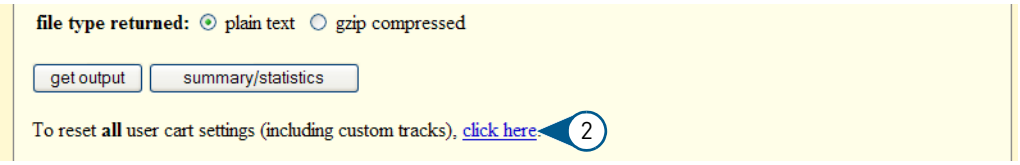



The Table Browser page appears as shown below with a description of the tool at the top of the page, the Table Browser tool in the middle, and guidelines for use at the bottom.




**2** **(Recommended) Reset the user cart settings**

(Recommended) In the Table Browser page, scroll to the bottom of the page, then click the **click here** link in the line of text at the bottom of the window (below the "get output" button) to reset all user cart settings.



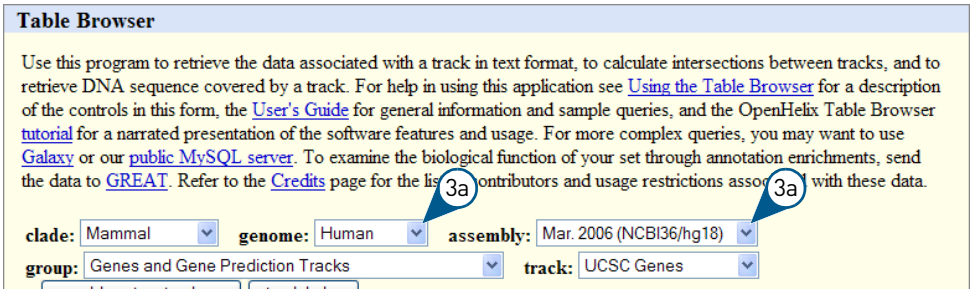

**Note:** The UCSC browser remembers user settings. Clicking the "click here" link ensures that you are using the correct settings rather than settings that might have been retained from a previous browsing session. For example, clicking the link also automatically resets the genome as "Human" and the assembly as "Feb. 2009". Note that the example still follows hg18 (Feb. 2006).

## 3 Configure the Table Browser settings

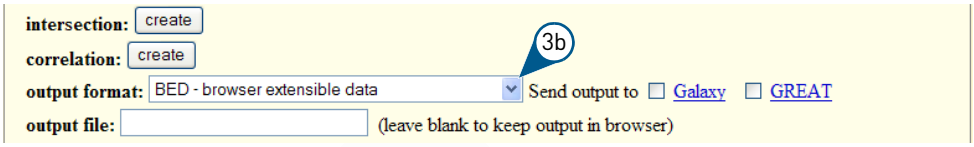
a. In the Table Browser page, specify the settings as desired.

For example, to use the March 2006 build of the human genome you would specify:

- genome – **Human**
- assembly – **Mar. 2006**

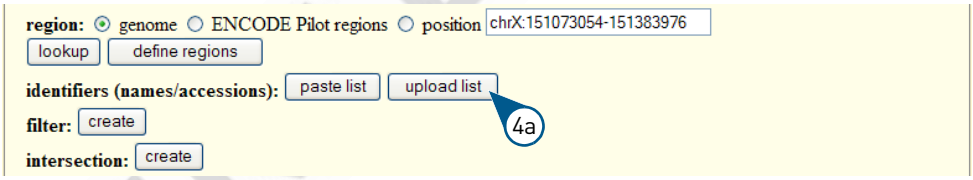



b. From the output format drop-down list, select **BED - browser extensible data**.



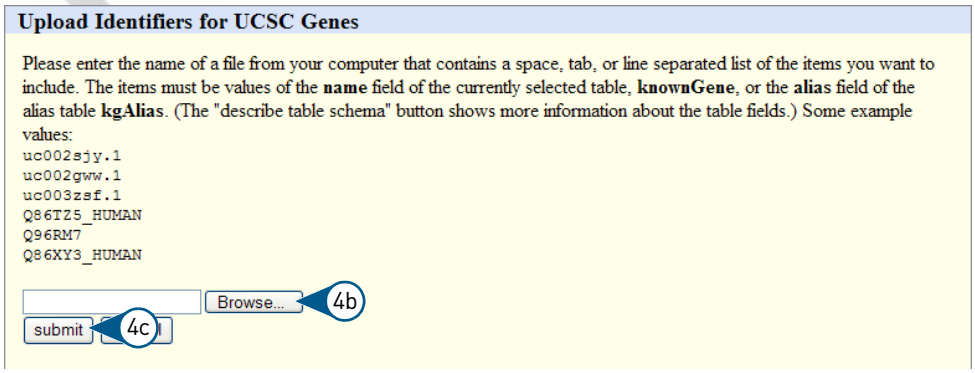

## 4 Upload the gene list file

a. In the Table Browser page, click **upload list** beside the identifiers (names/accessions) field.




b. In the Upload Identifiers for UCSC Genes page, click **Browse**, then select the tab-delimited text file containing your gene list that you created in step 2.

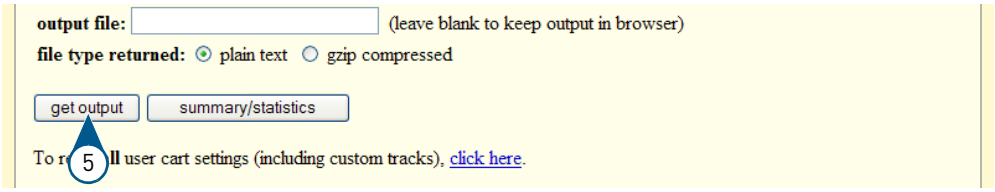
c. Click **submit**.




## 5 Generate the output

In the Table Browser page, click **get output**.

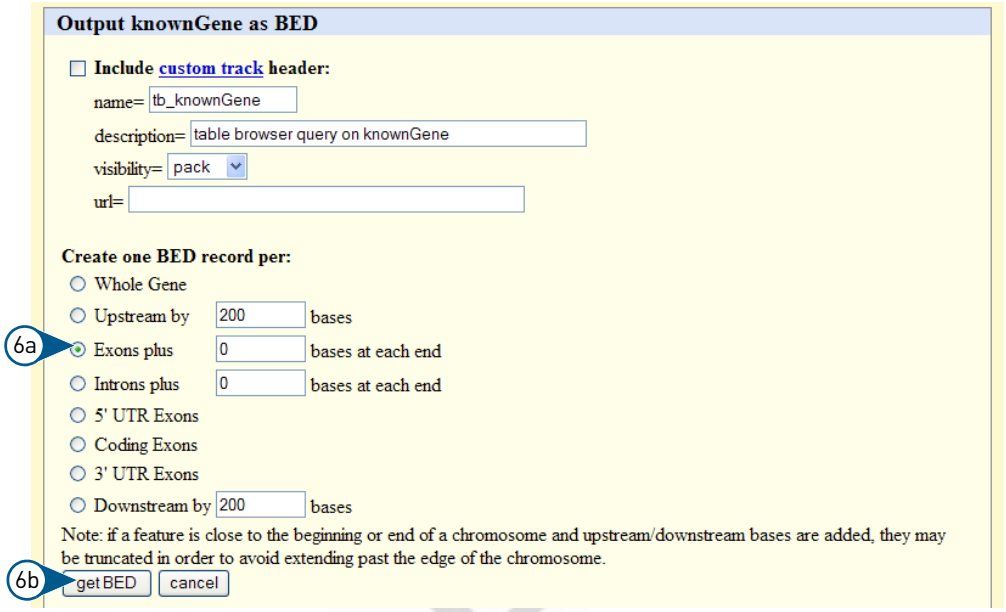
**IMPORTANT!** Do not change any other parameters.



## 6 Configure the settings for the output

a. In the Output known Gene as BED page, select **Exons plus 0 bases at each end**.

b. Click **get BED**.




The UCSC Genome Browser generates the output as shown in the example below.

```
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       -
```

## 7 Save the BED output to a data file

a. Select and copy all of the text displayed in the browser.

b. In Microsoft Excel, select **Edit ▸ Paste Special** to paste the output into a blank worksheet.

c. In the Paste Special dialog, select **Text** in the Paste As field, then click **OK**.

| | 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 | - | |

d. Delete all columns except Chromosome, Exon Start, and Exon Stop coordinates

For example, in the figure above, keep columns A, B, and C and remove the rest.

e. In the Save As dialog box, enter a name for the coordinates list, select **Text (Tab-delimited)(*.txt)** from the Save as type drop-down list, then click **Save**.


If Excel warns you that:

- Your output format does not support multiple worksheets, click **OK**.
- Tab-delimited text may not support some characters, click **Yes** to ensure that the file contains only text characters.

## 8 Submit the completed file

Email the saved tab-delimited text file (coordinates file) to Life Technologies along with your purchase order and the completed TargetSeq™ Custom Enrichment Kit Library Specification Form.

## Reviewing and approving your custom design

This section describes how to review and approve custom TargetSeq™ Custom Enrichment Kit designs sent to you by Life Technologies.

### About the design file formats

After Life Technologies accepts your coordinates file and completes a probe design, you will receive a set of design files in an email from the design team. Life Technologies provides design file(s) in the formats shown in the following table.

| File format | Extension | View using... |
|---|---|---|
| BED (Browser Extensible Data) | .bed | UCSC Genome Browser (website) |
| Summary Design | .txt | Microsoft WordPad or Notepad |

### Guidelines for reviewing your design files

Consider the following while reviewing your design:

- The .bed file contains the following tracks:

| Track | Description |
|---|---|
| primary_target_region | Lists the genomic regions that you provided. |
| capture_target | Lists regions actually covered by the design. |

- When reviewing the design file, focus on gaps in the capture_target track that do not appear in the primary_target_region track. The gaps represent portions of your target regions not covered by the design.

  **Note:** If two or more of your target regions overlap, Life Technologies will merge them into one region. In addition, regions that are less than 100bp will be padded to a minimum size of 100bp.

- Regions not covered by the design are usually repetitive regions that, if included, cause capture of other homologous regions in the genome and decrease capture efficiency. Therefore, most TargetSeq™ panel experiments benefit from excluding these regions from the design.
- The stringency filter Life Technologies uses by default will not include low-copy repeats (appearing 2-5 times in the genome) in the design. If these regions are necessary to your research, indicate this requirement in the email to Life Technologies (see "Step 3. Approve or request changes to the design" on page 11) and Life Technologies will use less stringent criteria in generating the design.

  **Note:** Less-stringent designs may provide more genomic coverage but at the cost of decrease capture efficiency, resulting in more off-target reads when the captured DNA is sequenced.

**Note:** If you have questions, contact Life Technologies Support (see "Obtaining support" on page 12).

## Step 1. Visually review the design

In this procedure, you will use the UCSC Genome Browser website to review the Browser Extensible Data (.bed) file that contains the probe design provided by Life Technologies. For more information on the UCSC Genome Browser website, go to **http://genome.ucsc.edu**.

**1 Save the BED file**

Save the BED file provided via email to your computer or a network drive.

**2 Open the UCSC Genome Browser Gateway**

a. Using a web browser, open the UCSC Genome Browser website (**http://genome.ucsc.edu**).

b. In the left pane of the UCSC Genome Browser website, click **Genome Browser**.

**3 Upload the custom tracks**

a. In the Genome Browser Gateway page, select the options appropriate for your design in the genome and assembly list boxes.

**Note:** For example, this procedure uses "Human" for the build and "Mar. 2006" as the assembly.

b. Click **add custom tracks**.

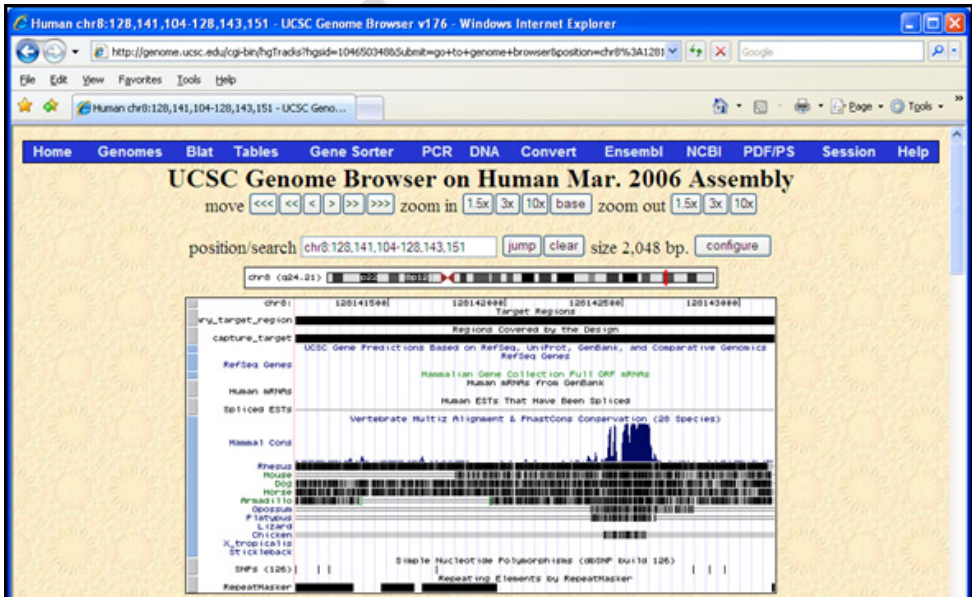
If the Gateway page displays the "manage custom tracks" button, click **Click here** to reset to display the "add custom tracks" button.

c. In the Add Custom Tracks page, click **Browse**, upload the BED file, then click **Submit**.

**4 Review the design**

a. In the Manage Custom Tracks page, click **go to genome browser** to view the regions.

The following figure shows an example design displayed in the UCSC Genome browser.




b. Review the design using the guidelines in "Guidelines for reviewing your design files" on page 9.

The UCSC Genome Browser provides the features in the table below to help you review the design. For more information on the UCSC Genome Browser features, click **Help**.

| To... | Action |
|---|---|
| Zoom the view | Click the **zoom in** and **zoom out** buttons to zoom in or out on the center of the annotation tracks window by 1.5-, 3-, or 10-fold. |
| Scroll the view | Click the move buttons to scroll to the left or right. |
| Display a different position in the genome | Enter coordinates in the **position/search** field, then click **jump**. |
| View the base composition | Click **base** to view the base composition of the sequence underlying the current annotation track display. |

## Step 2. Review the Summary Design File

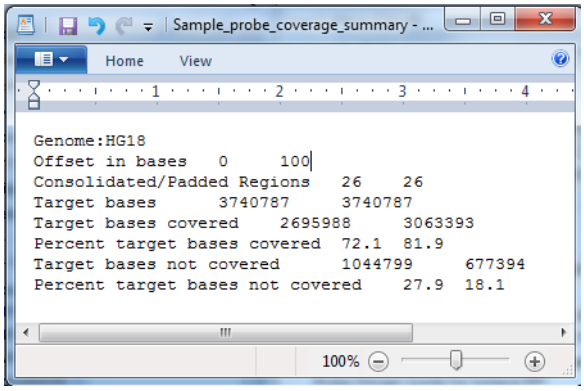
In this procedure, you will review the Summary Design File that contains the details of the custom design provided by Life Technologies.

**1** **Open the summary design file**

Open the summary design file (.txt) using a text editor, such as Microsoft WordPad, Microsoft Notepad, or Microsoft Excel.

**2** **Review the design to ensure that it meets your specifications**

Review each field to ensure that the design meets the specifications for your project. See the table below for more information on the summary design file fields.




| Field | Definition |
|---|---|
| Genome | The genome and build targeted by the design. |
| Offset in bases | The amount of additional padding added to both sides of probe to simulate additional coverage from sequence of captured fragments. |
| Consolidated/ Padded Regions | The number of regions in target file(s) after overlapping regions have been padded and merged. |
| Target bases | The number of bases in the merged target regions. |
| Target bases covered | The number of bases of merged targets that are covered by probes using the given offset. |
| Percent target bases covered | The percentage of merged targets that are covered by probes using the given offset. |
| Target bases not covered | The number of bases of merged targets that are not covered by any probe sequence using the given offset. |
| Percent target bases not covered | The percentage of merged targets that are not covered by any probe sequence using the given offset. |

## Step 3. Approve or request changes to the design

After you have reviewed the design files, you can:

- Approve the design
- Request changes to the design
- Request additional information from Life Technologies if you have questions on the design

For tracking purposes, reply to the email that included the design files with the proposed design. In your reply, either indicate that you approve the design or specify any changes needed for approval.

**IMPORTANT!** Life Technologies cannot finalize the design and produce the TargetSeq™ capture probe pools until we receive approval.

## Obtaining support

For the latest services and support information for all locations, go to:

**www.iontorrent.com/support/** or **www.lifetechnologies.com/support**

At the website, you can:

- Access worldwide telephone and fax numbers to contact Technical Support and Sales facilities
- Search through frequently asked questions (FAQs)
- Submit a question directly to Technical Support
- Search for user documents, application notes, and other product support documents



**For Research Use Only. Not intended for any animal or human therapeutic or diagnostic use.**



**WEB SITE**

**www.iontorrent.com**

**IMPORTANT PHONE NUMBERS**

If you are located in North America, please contact Ion Torrent at:

1-87-SEQUENCE (1-877-378-3623)

If you are located outside of North America, please contact Ion Torrent at:

+1-203-458-8552

**ION COMMUNITY**

**ioncommunity.iontorrent.com**

**SERVICE EMAIL**

**ionsupport@lifetech.com**

**ADDRESS**

Ion Torrent
246 Goose Lane
Suite 100
Guilford, CT 06437
USA

Ion Torrent
7000 Shoreline Court
Suite 201
South San Francisco, CA 94080
USA

Headquarters
5791 Van Allen Way | Carlsbad, CA 92008 USA | Phone +1 760 603 7200 | Toll Free in USA 800 955 6288
**For support visit** www.appliedbiosystems.com/support

www.lifetechnologies.com