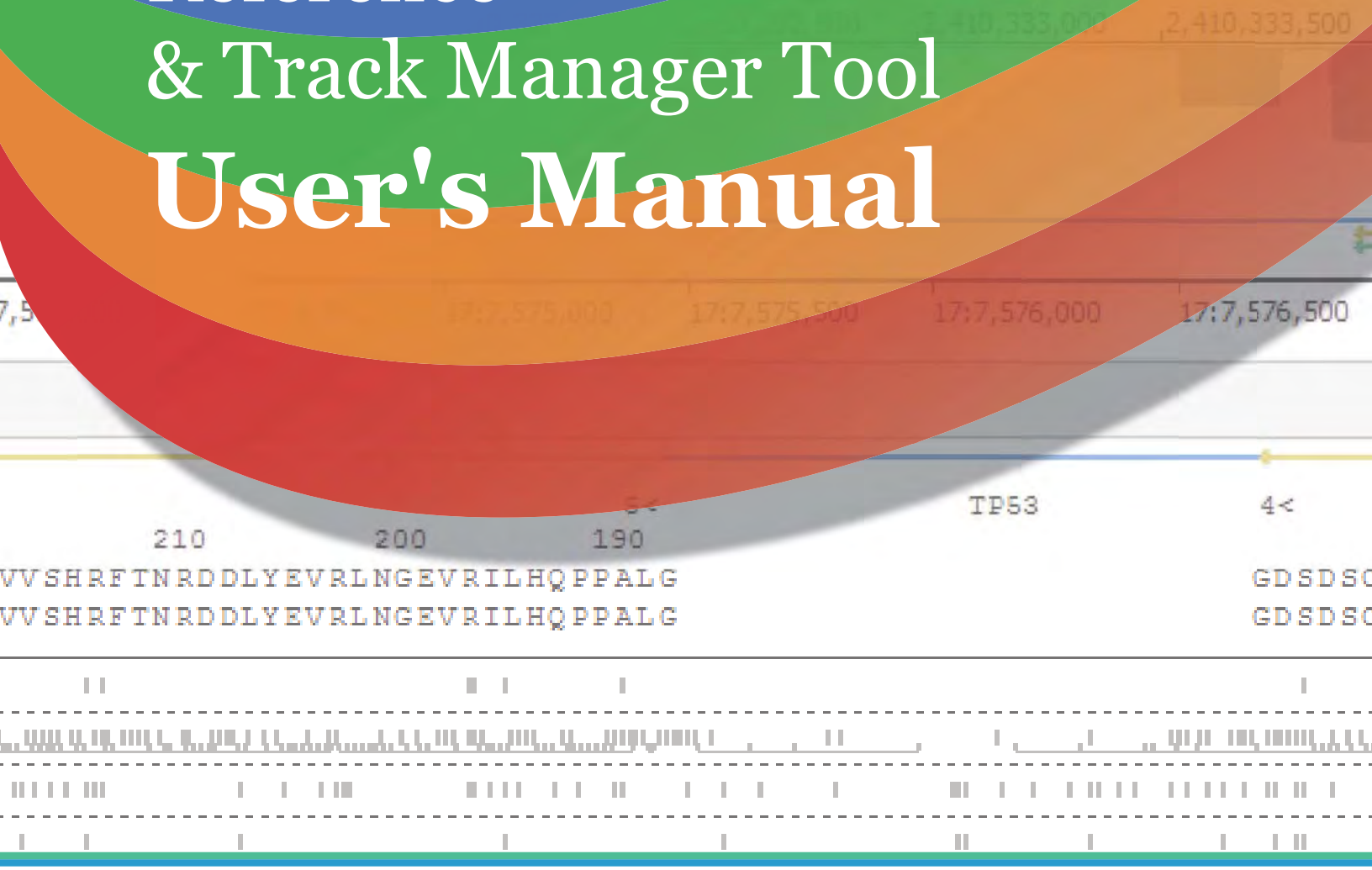
NextGENe®
2nd Generation Sequence Analysis Software

MUTATION Surveyor®
DNA Variant Analysis Software

Reference & Track Manager Tool User's Manual



SoftGenetics, LLC
100 Oakwood Avenue, Suite 350, State College, PA 16803 USA * info@softgenetics.com • www.softgenetics.com • 888-791-1270

Release Information

Document Version Number	*RefTrackMgr-1.0-UG001*
Software Version	*1.0*
Document Status	*Final*

Copyright

Limit of Liability

SoftGenetics, LLC has used their best effort in preparing this guide. SoftGenetics makes no representations or warranties with respect to the accuracy or completeness of the contents of this guide and specifically disclaims any implied warranties of merchantability or fitness for a particular purpose. Information in this document is subject to change without notice and does not represent a commitment on the part of SoftGenetics or any of its affiliates.The accuracy and completeness of the information contained herein and the opinions stated herein are not guaranteed or warranted to produce any particular results, and the advice and strategies contained herein may not be suitable for every user.

The software described herein is furnished under a license agreement or a non-disclosure agreement. The software may be copied or used only in accordance with the terms of the agreement. It is against the law to copy the software on any medium except as specifically allowed in the license or the non-disclosure agreement.

Trademarks

The name "*SoftGenetics*," the SoftGenetics logo, *NextGENe, Mutation Surveyor*, *Geneticist Assistant*, and the *SoftGenetics Reference and Track Manager Tool* are trademarks or registered trademarks of SoftGenetics, LLC. All other products and company names mentioned herein might be trademarks or registered trademarks of their respective owners.

Customer Support

Customer support is available to organizations that purchase NextGENe and that have an annual support agreement. Contact SoftGenetics at:

SoftGenetics, LLC
100 Oakwood Ave, Suite 350
State College, PA 16803
(814) 237-9340
(888) 791-1270 (US Only)
tech_support@softgenetics.com
www.softgenetics.com

Table of Contents

Preface

Welcome to the *Reference & Track Manager Tool User's Manual*. The purpose of the *Reference & Track Manager Tool User's Manual* is to answer your questions and guide you through the procedures necessary to use the NextGENe application efficiently and effectively.

Using the manual

You will find the *Reference & Track Manager Tool User's Manual* easy to use. You can simply look up the topic that you need in the table of contents or the index. Later, in this Preface, you will find a brief discussion of each chapter to further assist you in locating the information that you need.

Special information about the manual

The *Reference & Track Manager Tool User's Manual* has a dual purpose design. It can be distributed electronically and then printed on an as-needed basis, or it can be viewed online in its fully interactive capacity. If you print the document, for best results, it is recommended that you print it on a duplex printer; however, single-sided printing will also work. If you view the document online, a standard set of bookmarks appears in a frame on the left side of the document window for navigation through the document. For better viewing, decrease the size of the bookmark frame and use the magnification box to increase the magnification of the document to your viewing preference.

Conventions used in the manual

The *Reference & Track Manager Tool User's Manual* uses the following conventions:

- Information that can vary in a command—variable information—is indicated by alphanumeric characters enclosed in angle brackets; for example, <Project Name>. Do not type the angle brackets when you specify the variable information.
- A new term, or term that must be emphasized for clarity of procedures, is *italicized.*
- Page numbering is "online friendly." Pages are numbered from 1 to x, *starting with the cover* and ending on the last page of the index.

Although numbering begins on the cover page, this number is not visible on the cover page or front matter pages. Page numbers are visible beginning with the first page of the table of contents.

- This manual is intended for both print and online viewing.
    - If information appears in blue, it is a hyperlink. Table of Contents and Index entries are also hyperlinks. Click the hyperlink to advance to the referenced information.

Assumptions for the manual

The *Reference & Track Manager Tool User's Manual* assumes that:

- You are familiar with windows-based applications and basic Windows functions and navigational elements.
- References to any third party standards or third party software functions were current as of the release of this version of the tool, and might have already changed.

Organization of the manual

In addition to this Preface, the *Reference & Track Manager Tool User's Manual* contains the following chapters:

- Chapter 1, "Getting Started with the Reference & Track Manager Tool," on page 5 explains how to launch the Reference and Track Manager Tool and details the layout of the tool.
- Chapter 2, "Using the Reference & Track Manager Tool," on page 9 details all the functions that are available from the Reference & Track Manager tool to import preloaded references and data from any public or proprietary variant database into NextGENe or Mutation Surveyor as well as all the functions that are available for managing the references and data after importing.

Chapter 1
Getting Started with the Reference & Track Manager Tool

You use the Reference & Track Manager tool to import and manage preloaded references and data from any public or proprietary variant database into NextGENe or Mutation Surveyor. The imported variant database is called a *track*. This chapter explains how to launch the Reference & Track Manager tool and details the layout of the tool.

This chapter covers the following topics:

Opening the Reference & Track Manager Tool

When you install NextGENe or Mutation Surveyor, the Reference & Track Manager tool is also installed. You can open the tool directly from NextGENe or Mutation Surveyor, or you can open the tool as a standalone application.

After you open the Reference &Track Manager tool, your current instance of SoftGenetics software (NextGENe or Mutation Surveyor) is unavailable. You must close the Reference & Track Manager tool to continue with work in NextGENe or Mutation Surveyor.

To open the Reference & Track Manager tool

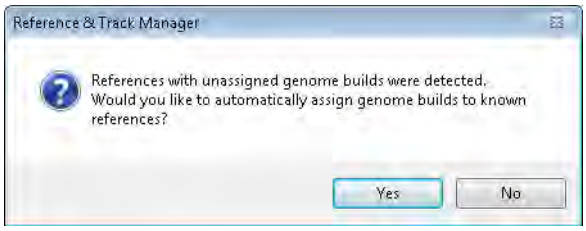
1. Do any of the following as appropriate based on the product that you are using to access the Reference & Track Manager tool

Product	Options
NextGENe	• On the NextGENe main menu, click Tools > Reference & Track Manager. • On the Start menu, select All Programs\SoftGenetics\NextGENe\RTManager.
Mutation Surveyor	• On the Mutation Surveyor main menu, click Tools > Reference & Track Manager. • On the Start menu, select All Programs\SoftGenetics\Mutation Surveyor\RTManager.

Two options are possible:

- The Reference & Track Manager dialog box opens.
- If any of the preloaded references do not have assigned genome builds, then a message opens indicating this, and prompting you to automatically assign the builds. Click Yes to automatically assign the genome builds, or click No to prevent the automatic assignment and open the Reference & Track Manager tool.

Figure 1-1: No Assigned Genome Builds message

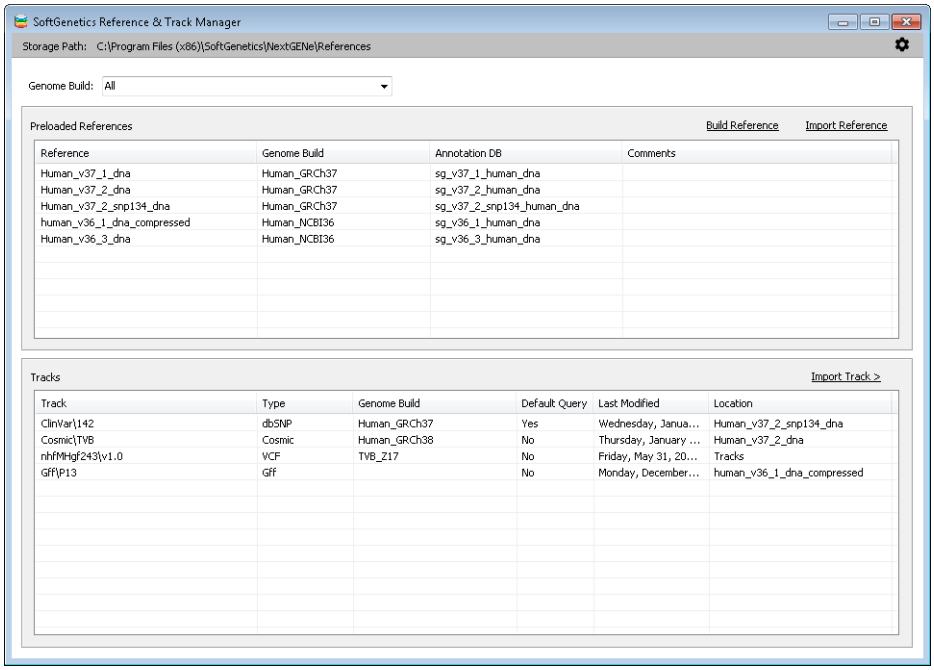


You can always automatically or manually assign the genome builds at a later date. See "Importing and Managing Preloaded References and Variant Databases with the Reference & Track Manager Tool" on page 11.

2. Continue to "The Reference & Track Manager Tool Layout" on page 7.

The Reference & Track Manager Tool Layout

Figure 1-2: Reference & Track Manager Tool dialog box



From top to bottom, the Reference & Track Manager tool dialog box displays the following options and information:

- The Storage Path, which is the directory that has been set for preloaded references.
- A Genome Build dropdown list, from which you can select the appropriate genome build for preloaded references and tracks.
- The Preloaded References pane, which displays all previously imported preloaded references. Options are also available in this pane for building references and importing references.

The option to build a reference is enabled only for NextGENe. Although it is also displayed for Mutation Surveyor, it is not enabled.

- The Tracks pane, which displays any previously imported databases, custom variant tracks, and/or gene annotation tracks.

Other notable information about the dialog box includes the following:

- In the Tracks pane, the Default Query status indicates whether the track, by default, is queried for all projects that are aligned to a reference with the genome build version that is specified for the track.
- In the Genome Build field of the Preloaded References pane, the reference name, which

includes the genome build, is displayed for any preloaded reference that was imported prior Version 1.0 of the Reference & Track Manager tool. Beginning with Version 1.0 of Reference & Track Manager tool, when you import any preloaded reference, the Genome Build for the imported reference is displayed.

- Prior to Version 1.0 of Reference & Track Manager tool, when you imported any track, the data was stored directly in the selected Reference folder, and the location of the reference folder is displayed in the Location column in the Tracks pane to indicate this. Beginning with Version 1.0 of Reference & Track Manager tool, when you import any track, the data is stored in a Tracks folder in the Reference directory, and “Tracks” is displayed in the Location column in the Tracks pane to indicate this.

Chapter 2
Using the Reference & Track Manager Tool

You use the Reference & Track Manager tool to import preloaded references and data from any public or proprietary variant database into NextGENe or Mutation Surveyor. The imported variant database is called a *track*. You can import variant information with clinical significance values from the ClinVar database, coding and non-coding variant information from the COSMIC database, functional prediction scores, population frequencies, clinical significance and ensemble prediction scores from the dbNSFP database, and ensemble prediction scores for splice variants from the dbscSNV database. You can also use the tool to import custom variation databases and gene annotation tracks into NextGENe or Mutation Surveyor, and to load track data for previously run projects. You can also use the tool's options to manage the imported references and tracks, such as setting the default query status for a track.

This chapter covers the following topics:

Chapter 2
Using the Reference & Track Manager Tool

Importing and Managing Preloaded References and Variant Databases with the Reference & Track Manager Tool

You use the Reference & Track Manager tool to import preloaded references and data from any public or proprietary variant database into NextGENe or Mutation Surveyor. The imported variant database is called a *track*. You can also use the tool's options to manage the imported references and tracks, such as setting the default query status for a track.

To import and manage preloaded references and variant databases with the Reference & Track Manager tool

To open the Reference & Track Manager tool, see "Opening the Reference & Track Manager Tool" on page 6.

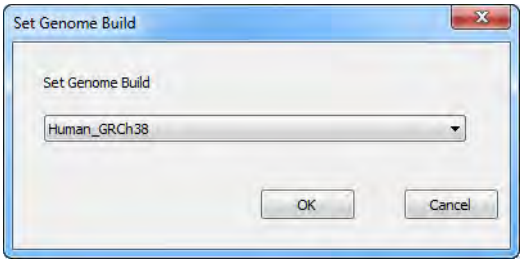
Optionally, you can do any or all the following as needed:

- To filter the display to only those references and tracks that used a specific genome build, select the appropriate build on the Genome Build dropdown list.
- To edit the Default Query status for a track, right-click the track, and on the context menu that opens, click Default Query, and then select Yes or No as appropriate.

The default value for the Default Query status is Yes, which allows NextGENe or Mutation Surveyor to automatically query a track, for example, the dbSNP track, for alignments to a reference of the genome build that is specified for the track.

- To delete a reference or track, right-click the reference or track, and on the context menu that opens, click Delete Reference or Delete Track as appropriate.
- To manually set the genome build for any reference or track that was imported prior to Version 1.0 of Reference & Track Manager tool, right-click the reference or track (CTRL-click to select multiple references or tracks), and on the context menu that opens, click Set Genome Build, and then in the Set Genome Build dialog box, select the appropriate build.

Figure 2-1: Set Genome Build dialog box

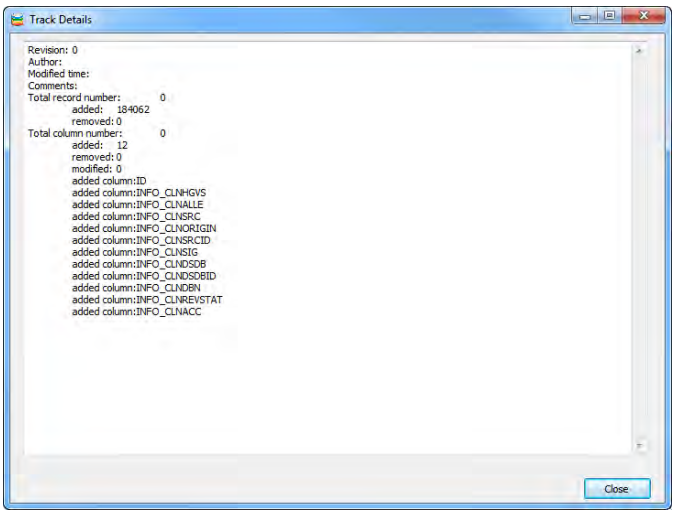


- To use some of the options on the Settings menu, click the Settings icon , and on the Settings menu that opens, do any or all of the following as needed:

Action	Step
To specify a different directory for preloaded references	See "To specify a directory for preloaded reference (Set Storage Path)" on page 13.
To edit genome build tags	See "To edit genome build tags" on page 14.
To automatically assign genome builds to preloaded references	Click Tools > Assign Genome Builds to Known References.
To move references to the current reference storage site	See "To move references to the current reference storage site" on page 15.
To work with User Management settings	Click User Management. **Note:** For detailed information about User Management, see the *NextGENe User's Manual or Online Help*, or the *Mutation Surveyor User's Manual or Online Help.*
To open the Online Help that is specific for the Reference & Track Manager tool	Click Help.
To open the About dialog box that displays information about the Reference & Track Manager tool, such as the version that you are currently using	Click About.

- To view the detailed information for a track, right-click the track, and on the context menu that opens, click Show Detail to open the Track Details dialog box. The dialog box shows the date that the track was imported, the file that was used for the track, and the columns that are included in the track.

Figure 2-2: Track Details dialog box



- To import data from the dbSNP or ClinVar database, continue to "To import data from the dbSNP or ClinVar database" on page 16.
- To import data from the COSMIC database, continue to "To import data from the COSMIC database" on page 17.
- To import data from the dbNSFP database, continue to "To import data from the dbNSFP database" on page 18.
- To import data from the dbscSNV database, continue to "To import data from the dbscSNV database" on page 21.
- To import data from other custom variation databases, continue to "To import data from other variation databases" on page 22.
- To import gene annotation tracks, continue to "To import gene annotation tracks" on page 23.
- To import a preloaded reference, continue to "Importing Preloaded References" on page 26.
- To build a preloaded reference, click Build Reference, and then continue to "The Build Preloaded Reference Tool (For NextGENe Only)" on page 32.

To specify a directory for preloaded reference (Set Storage Path)

By default, the directory for preloaded references is:

C:\Users\Public\Documents\SoftGenetics\References

You can leave this value as-is, or you can select a different location from the Settings menu for the Reference & Track Manager tool.

The directory that you specify here for preloaded references also sets the directory for the Build Preloaded Reference tool (see "The Build Preloaded Reference Tool (For NextGENe Only)" on page 32) and the directory for preloaded references that you import into NextGENe or Mutation Surveyor. (See "Importing Preloaded References" on page 26.)

1. Click the Settings icon  , and on the Settings menu that opens, click Set Storage Path to open the Browse for Folder dialog box.
2. In the Browse for Folder dialog box, browse to and select an existing folder in which the preloaded reference files are to be stored, or make a new folder, and then click OK.

   The Browse for Folder dialog box closes. The Reference & Track Manager tool remains open. The newly selected storage path is displayed in the title bar of the dialog box.

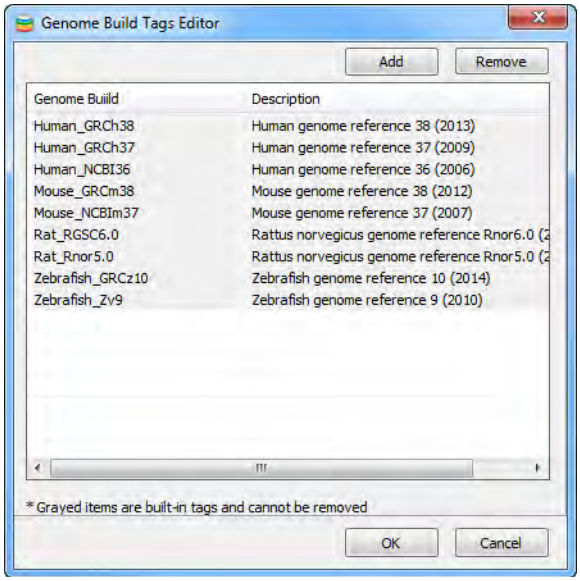
To edit genome build tags

Built-in genome tags are supplied for the Reference & Track Manager tool. You cannot edit or delete these built-in tags but you can add, edit, or remove custom genome build tags.

1. Click the Settings icon , and on the Settings menu that opens, click Edit Genome Build Tags.

   The Genome Build Tags Editor dialog box opens. A list of all the built-in genome tags is displayed in the dialog box. The tags are highlighted in gray.

Figure 2-3: Genome Build Tags Editor dialog box



2. Do any or all the following as needed:

   - To add a new genome build tag, click Add, and then double-click in the blank line that is displayed to enter the name and description for the genome build tag.
   - To edit a custom genome build tag, double-click to select the tag name and/or description, and then edit the information as appropriate.
   - To delete a genome build tag, select the tag, and then click Remove.

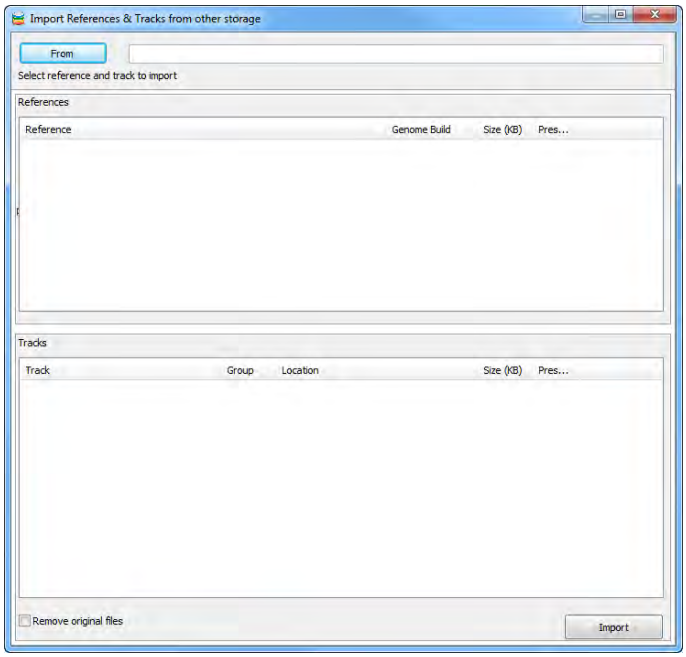
Deleting a genome build tag does not delete it for any previously imported databases or tracks. Going forward, the tag is simply not available for selection.

To move references to the current reference storage site

If you have references that are stored in a different directory than the directory that is currently specified for preloaded references, you can move these references to the current references storage site.

For information about setting the reference storage site, see "To specify a directory for preloaded reference (Set Storage Path)" on page 13.

Figure 2-4: Import References & Tracks from other storage dialog box



1. If you have not done so already, click the Settings icon, and on the Settings menu that opens, click Tools > Import References & Tracks from other storage.

   The Import References & Tracks from other storage dialog box opens.

2. Click From to open the Browse for Folder dialog box.

3. In the Browse for Folder dialog box, browse to and select the folder that contains the references that are to be moved, and then click OK.

   All the references that are contained in the selected folder are displayed in the References pane. If tracks are associated with a selected reference, then these tracks are displayed in the Tracks pane.

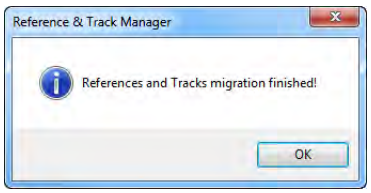
4. Optionally, to delete the original source files from the alternate storage site after moving them, select Remove Original Files.

5. Click Import.

   A copy of each reference that is to be moved is made, and then this copy is imported into the current reference storage site.

   After all the references are successfully imported, a message opens indicating this.

Figure 2-5: References & Tracks successfully imported message



6. Click OK to close the message. The Import References & Tracks from other storage dialog box remain open.

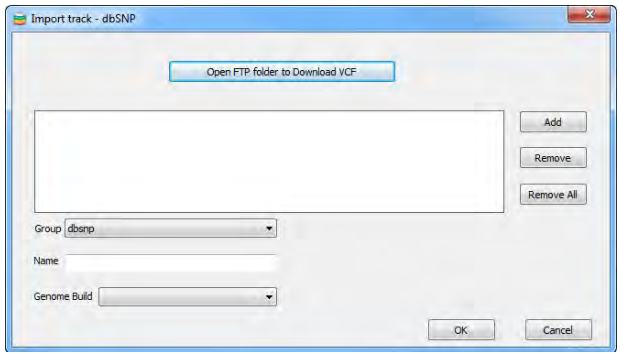
To import data from the dbSNP or ClinVar database

You can import data from the ClinVar or other dbSNP databases from NCBI. When you import a ClinVar database, the clinical significance and clinical origin for each variant are also automatically imported.

1. In the Tracks pane, click Import Track, and on the dropdown menu that opens, click dbSNP/ClinVar.

   The Import track - dbSNP dialog box opens.

Figure 2-6: Import track - dbSNP dialog box



2. Choose the appropriate group—dbsnp or clinvar.
3. Click Open FTP Folder to Download VCF.

   The NCBI FTP site opens. This site contains all the dbSNP or ClinVar database files that are available for downloading.

4. Download the appropriate version of the database.
5. Click Add to browse to and select the downloaded files.

6. In the Name field, enter the name or version number for the downloaded database.
7. On the Genome Build dropdown list, select the genome build for the track.
8. Click OK.

   After the import is complete, an Import Complete message is displayed, and the Import track - dbSNP dialog box closes. An entry for the imported database is displayed in the Tracks pane.
9. Optionally, to set the Default Query status to No for the database, right-click the track name in the Tracks pane, and on the context menu that opens, select Default Query > No.

If the Default Query status is set to No, then NextGENe or Mutation Surveyor cannot automatically query the dbSNP or ClinVar track for alignments to a reference of the genome build that is specified for the track.

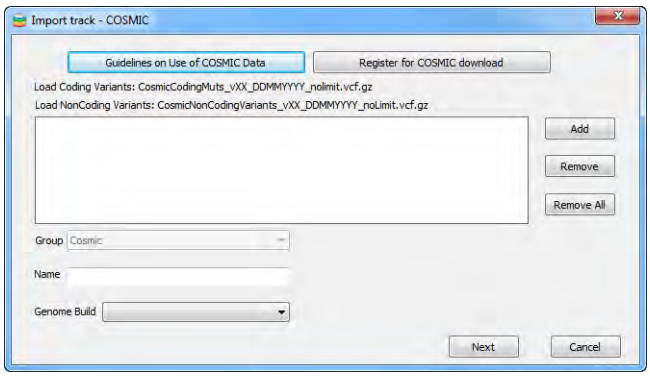
To load dbSNP or ClinVar information for previously run projects, continue to "To load track data for previously run projects" on page 24.

To import data from the COSMIC database

1. In the Tracks pane, click Import Track, and on the dropdown menu that opens, click COSMIC.

   The Import track - COSMIC dialog box opens.

Figure 2-7: Import track - COSMIC dialog box



Optionally, click Guidelines on Use of COSMIC data to go to a web page provided by Sanger with guidelines and information about the public use of COSMIC data.

2. To register for COSMIC downloads, click Register for COSMIC download.

   The COSMIC database registration page opens.
3. After correctly registering for COSMIC downloads, follow the instructions from COSMIC to download VCF files for the database.

The exact file name changes with new versions of the database.

4. Click Add to browse to and select the downloaded files.
5. In the Name field, enter the name or version number for the downloaded files.
6. On the Genome Build dropdown list, select the genome build for the track.
7. Click OK.

   After the import is complete, an Import Complete message is displayed, and the Import track - COSMIC dialog box closes. An entry for the imported database is displayed in the Tracks pane.
8. Optionally, to set the Default Query status to No for the database, right-click the track name in the Tracks pane, and on the context menu that opens, select Default Query > No.

If the Default Query status is set to No, then NextGENe or Mutation Surveyor cannot automatically query the COSMIC track for alignments to a reference of the genome build that is specified for the track.

To load COSMIC tags for previously run projects, continue *"To load track data for previously run projects" on page 24.*

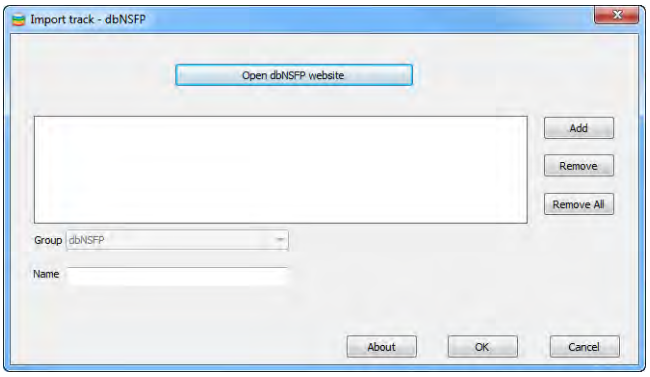
To import data from the dbNSFP database

A dbNSFP database file contains coordinates for multiple genome builds. As a result, after a dbNSFP database file is imported and extracted, a dialog box opens that displays all the compatible genome builds for the file. You must select the appropriate genome build from this list of compatible builds, and you can select multiple genome builds as needed.

1. In the Tracks pane, click Import Track, and on the dropdown menu that opens, click dbNSFP.

   The Import track - dbNSFP dialog box opens.

Figure 2-8: Import track - dbNSFP dialog box



Optionally, click About to open a dialog box that provides a link to an article that details the dbNSFP database.

2. Click Open dbNSFP website.

   The dbNSFP website page opens.

3. Download the appropriate version of the database for your work.

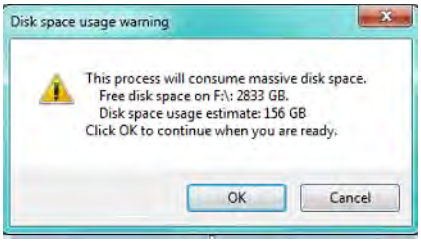
The latest dbNFSP version that NextGENe currently supports is v2.9.

4. Click Add to browse to and select the downloaded files.
5. In the Name field, enter the name or version number for the downloaded database.
6. Click OK.

   Two results are possible:

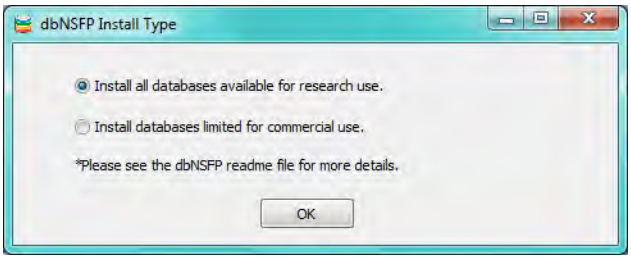
   - A Disk space usage warning message might open informing you that importing and extracting the selected dbNSFP database files will consume a large amount of disk space, the amount of disk space that will be required to extract the files, and the amount of free disk space that is currently available on the NextGENe client. Continue to Step 7.

Figure 2-9: Disk space usage warning message



   - The dbNSFP Install Type dialog box opens. By default, the option to install all databases that are available for research use is selected. Continue to Step 8.

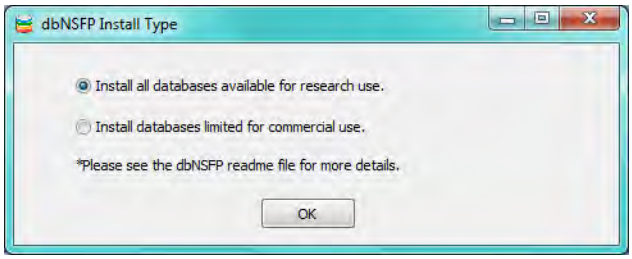
Figure 2-10: dbNSFP Install Type dialog box



7. If sufficient disk space is available, then click OK; otherwise, free up the appropriate amount of disk space for the client, and then click OK.

   The dbNSFP Install Type dialog box opens. By default, the option to install all databases that are available for research use is selected.

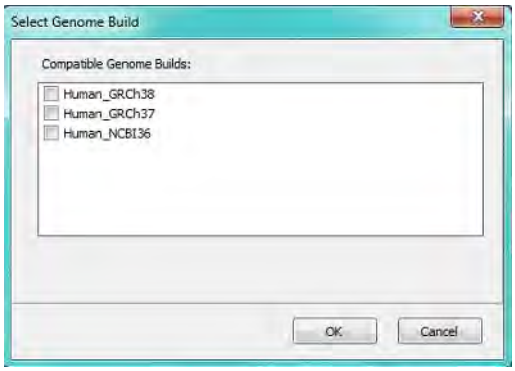
Figure 2-11: The dbNSFP Install Type dialog box



8. Leave the database type set to the default value of all available research databases, or select Install databases limited for commercial use as appropriate, and then click OK.

   The dbNSFP Install Type dialog box closes and the files are extracted. After the files are extracted, the Select Genome Build dialog box opens.

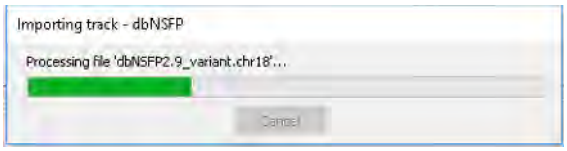
Figure 2-12: Select Genome Build dialog box



9. Select the appropriate compatible genome build, and then click OK.

   The Set Genome Build dialog box closes, and an Importing track - dbNSFP dialog box opens. The dialog box displays the progress of importing the selected database.

Figure 2-13: Importing track - dbNSFP dialog box



After the import is complete, Importing track - dbNSFP dialog box closes, and an Import Completed message is displayed. See Figure 2-14 on page 21.

Figure 2-14: Import Completed message

10. Click OK.

    The Import Completed message closes, and an entry for the imported database is displayed in the Tracks pane.

11. Optionally, to set the Default Query status to No for the database, right-click the track name in the Tracks pane, and on the context menu that opens, select Default Query > No.

If the Default Query status is set to No, then NextGENe or Mutation Surveyor cannot automatically query the dbNSFP track for alignments to a reference of the genome build that is specified for the track.

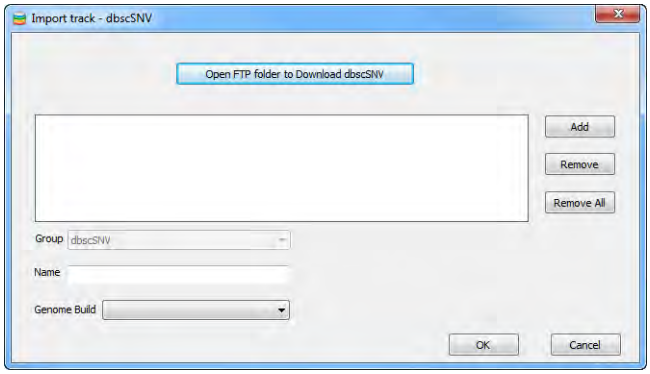
To load dbNSFP information for previously run projects, continue to "To load track data for previously run projects" on page 24.

To import data from the dbscSNV database

1. In the Tracks pane, click Import Track, and on the dropdown menu that opens, click dbscSNV.

   The Import dbscSNV dialog box opens.

Figure 2-15: Import dbscSNV dialog box



2. Click Open FTP folder to Download dbscSNV.

   A dbNSFP website page that has options for downloading the database opens.

3. Download the appropriate version of the database for your work.

The dbscSNV database is a database of all potential human SNVs within splicing consensus regions. It is listed as an Attached Database on the dbSNFP website.

---

4. Click Add to browse to and select the downloaded files.
5. In the Name field, enter the name or version number for the downloaded database.
6. On the Genome Build dropdown list, select the genome build for the track.
7. Click OK.

   After the import is complete, an Import Complete message is displayed, and the Import track - dbscSNV dialog box closes. An entry for the imported database is displayed in the Tracks pane.

8. Optionally, to set the Default Query status to No for the database, right-click the track name in the Tracks pane, and on the context menu that opens, select Default Query > No.

If the Default Query status is set to No, then NextGENe or Mutation Surveyor cannot automatically query the dbscSNV track for alignments to a reference of the genome build that is specified for the track.

To load dbscSNV information for previously run projects, continue to "To load track data for previously run projects" on page 24.

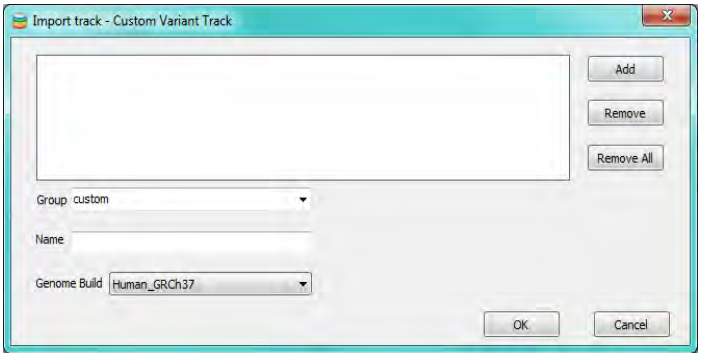
To import data from other variation databases

If you download data from variation databases other than dbSNP, ClinVar, COSMIC, dbNSFP, or dbscSNV, then you can also import this data into NextGENe.

1. In the Tracks pane, click Import Track, and on the dropdown menu that opens, click Custom Variant Track.

   The Import track - Custom Variant Track dialog box opens.

Figure 2-16: Import track - Custom Variant Track dialog box



2. Click Add to browse to and select the downloaded files.
3. In the Group field, leave the default value set to custom, or optionally, highlight custom, and then edit to reflect the custom track that you are importing.
4. In the Name field, enter the name or version number for the downloaded database.
5. On the Genome Build dropdown list, select the genome build for the track.
6. Click OK.

   After the import is complete, an Import Complete message is displayed, and the Import track - Custom Variant Track dialog box closes. An entry for the imported database is displayed in the Tracks pane.
7. Optionally, to set the Default Query status to No for the database, right-click the track name in the Tracks pane, and on the context menu that opens, select Default Query > No.

If the Default Query status is set to No, then NextGENe or Mutation Surveyor cannot automatically query the custom track for alignments to a reference of the genome build that is specified for the track.

To load variation information for previously run projects, continue to "To load track data for previously run projects" on page 24.

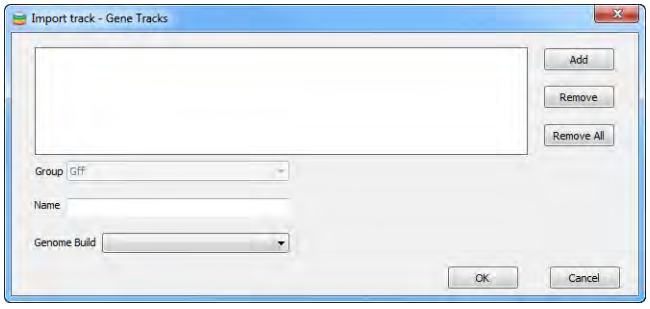
To import gene annotation tracks

You can import gene tracks from a file that is in either a .gff format or a .gff3 format. You can use this function to customize gene-level annotations such as gene names and transcripts.

1. In the Tracks pane, click Import Track, and on the dropdown menu that opens, click Gene Annotation Track.

   The Import track - Gene Tracks dialog box opens.

Figure 2-17: Import track - Gene Tracks dialog box



2. Click Add to browse to and select the downloaded files.
3. In the Name field, enter the name or version number for the downloaded database.
4. On the Genome Build dropdown list, select the genome build for the track.

5. Click OK.

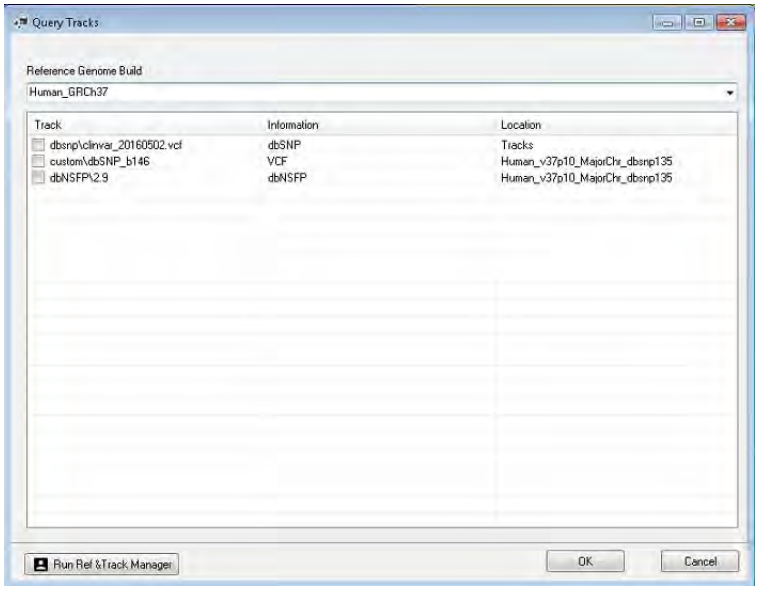
   After the import is complete, an Import Complete message is displayed, and the Import Gene Tracks dialog box closes. An entry for the imported track is displayed in the Tracks pane.

To load track data for previously run projects

1. Do one of the following:
   - In NextGENe, do the following:
     i Load the project in the NextGENe Viewer.
     ii. On the Viewer main menu, click Process > Query Reference Tracks.
   - In Mutation Surveyor, do the following:
     i Load the project.
     ii. On the main menu, click Process > Query Reference Tracks.

   The Query Tracks dialog box opens. The dialog box lists all the tracks that are available for the project reference.

Figure 2-18: Query Tracks dialog box



2. Select the reference genome build for the project.

   All the tracks that are available for the selected genome build are displayed in the dialog box. Tracks that have previously been queried for the project are selected.

3. Select the tracks that are to be queried for the project, or clear the selections for the tracks that are not to be queried.

4. Optionally, if the track that is to be queried for the project is not available, then click Run Ref & Track Manager to open the Reference & Track Manager tool and import the database.
5. Click Run Query.

   The Query Tracks dialog box closes. The Variation Tracks Settings dialog box opens.
6. Specify the filter and display settings for the track.

For detailed information about the Variation Tracks Settings dialog box, see the NextGENe User Manual or Online Help, or the Mutation Surveyor User's Manual or Online Help as appropriate.

The track information for the project is modified accordingly and the settings applied. If new tracks have been added to the project, then the tracks are loaded and the information from the tracks can be displayed in mutation reports in NextGENe or Mutation Surveyor.

Importing Preloaded References

If you are using the Reference & Track Manager tool with NextGENe, then to align to a reference larger than 250 Mbp for any application type, or to align to any reference sequence for the Transcriptome with Alternative Splicing application type, then you must do one of the following:

- Create a preloaded reference using the Build Preloaded Reference tool. (See "The Build Preloaded Reference Tool (For NextGENe Only)" on page 32.)
- Align the data against a preloaded reference that SoftGenetics supplies, either through the SoftGenetics ftp site, or on a DVD. For access to a needed reference, you have two options:
  - You can download and import preloaded references through the SoftGenetics ftp server.
  - You can import a preloaded reference from a DVD that SoftGenetics can send to you upon request.

If you are using the Reference & Track Manager tool with Mutation Surveyor, then you can use only human preloaded references. To obtain the appropriate human preloaded reference, you can do one of the following:

- Download and import preloaded references through the SoftGenetics ftp server.
- Import a preloaded reference from a DVD that SoftGenetics can send to you upon request.

After you import all your needed references, you can select the appropriate reference when you are aligning your data.

If you intend to import preloaded references in NextGENe, note the following:

- See http://www.softgenetics.com/NextGENe_011.html for a list of preloaded references that are available on the SoftGenetics ftp server and upon request on a DVD.
- If the genome you are interested in aligning to is not available, you can contact SoftGenetics and request a custom genome, or you can use the Build Preloaded Reference tool to create a preloaded reference. See "The Build Preloaded Reference Tool (For NextGENe Only)" on page 32.

For either NextGENe or Mutation Surveyor, you cannot import and use preloaded reference files if you have *not* installed MySQL. If you did not install MySQL when you installed NextGENe or Mutation Surveyor, then you can use the Reference Setup Wizard to do so.

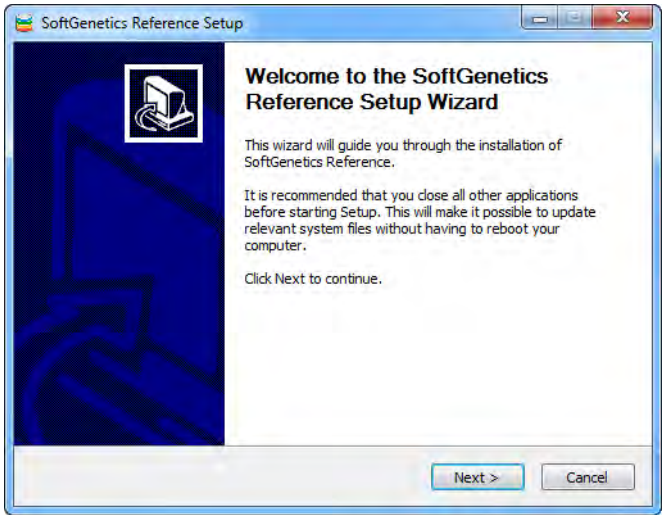
To import preloaded references

If you are importing a preloaded reference from a DVD, then make sure to insert the DVD into the client DVD/CD drive before you begin this procedure. You can repeat this procedure in its entirety as many times as needed to download and import all your required preloaded references.

1. Click Import Reference.

   The SoftGenetics Reference Setup Wizard opens.

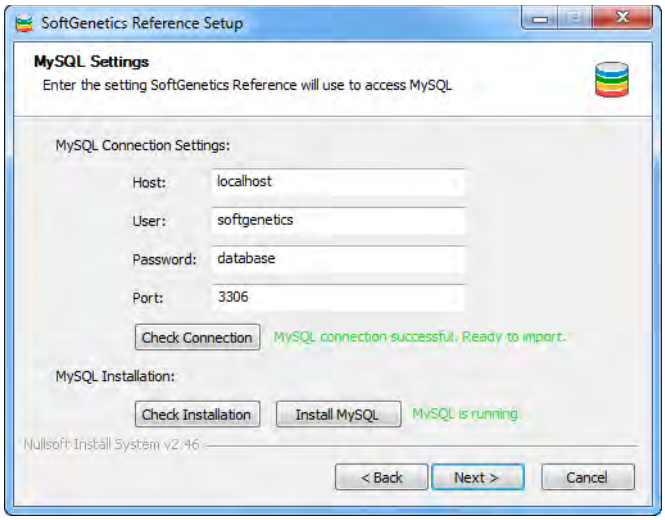
Figure 2-19: Reference Setup Wizard



2. Click Next.

   The MySQL Settings page opens. If MySQL has been installed correctly, and the connection to the database is successful, then “MySQL connection successful. Ready to Import” and “MySQL Installed” are displayed on the page, and you can continue to Step 4; otherwise, if either or both of these messages are not displayed, then continue to Step 3. See Figure 2-20 on page 28.

Figure 2-20: Reference Setup Wizard, MySQL Settings page



3. Do one or both of the following:

   - If “MySQL Installed” is not displayed on the page, then click Install MySQL.

     If MySQL cannot be installed successfully, contact tech_support@softgenetics.com.

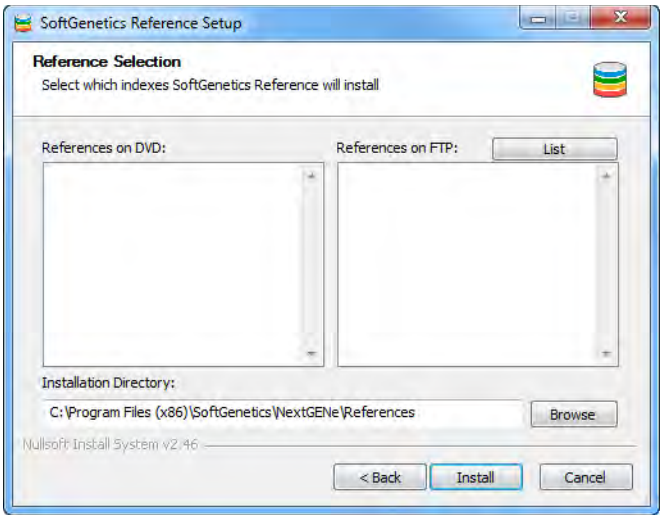
   - If “MySQL installed” is displayed, but “MySQL connection successful. Ready to Import” is *not* displayed, then click Check Connection.

     If the message MySQL Connection Successful is displayed, then continue to Step 4; otherwise, contact tech_support@softgenetics.com.

4. Click Next.

   The Reference Selection page opens. If you have inserted a DVD into the client DVD/CD drive, then the reference file that is on the DVD is listed in the References on DVD pane. See Figure 2-21 on page 29.

Figure 2-21: Reference Setup Wizard, Reference Selection page



5. If you are downloading a preloaded reference from the SoftGenetics ftp site, continue to Step 6; otherwise, if you are importing a preloaded reference from a DVD, then click the appropriate reference to select it, and then continue to Step 7.
6. To view all the available reference genomes on the SoftGenetics ftp server, click List.

   The References on FTP pane is populated with a list of all the available reference genomes.

Use the genomes that are appended with "_SOLID" or "_CS" strictly for SOLiD System data. Use all other genomes for Illumina, Roche, or Ion Torrent data. If the genome that you want to import is not available, you can contact SoftGenetics and request a custom genome or you can use Build Preloaded Reference tool to build a preloaded reference. See "The Build Preloaded Reference Tool (For NextGENe Only)" on page 32.

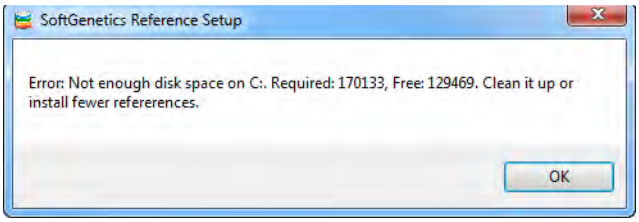
7. Select the reference file that is to be imported, or CTRL-click to select multiple non-continuous reference files, or SHIFT-click to select multiple continuous reference files.
8. The default installation directory for the preloaded reference files is C:\Users\Public\Documents\SoftGenetics\References. If you did not change this value in the Reference and Track Manager tool (see "To specify a directory for preloaded reference (Set Storage Path)" on page 13), then leave it as-is; otherwise, change the value.

9. Click Install.

   A Reference Setup message opens, indicating one of the following:

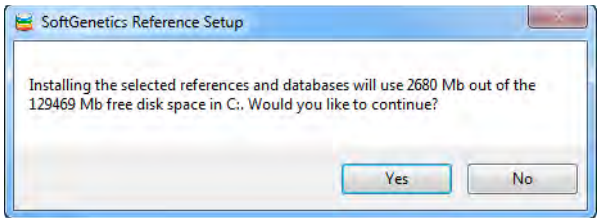
   - That there is not enough disk space to install the selected references and the required disk space for installation. It also prompts you to clean up the installation directory or install fewer references. Go to Step 10.

Figure 2-22: Reference Setup message, insufficient disk space



   - The amount of space (in Mb) that the installation of the selected references and databases will use, the amount of free disk space at the top level of the installation directory, and asking you if you want to continue. Go to Step 11.

Figure 2-23: Reference Setup message, sufficient disk space



10. Click OK to close the message, and to one of the following:

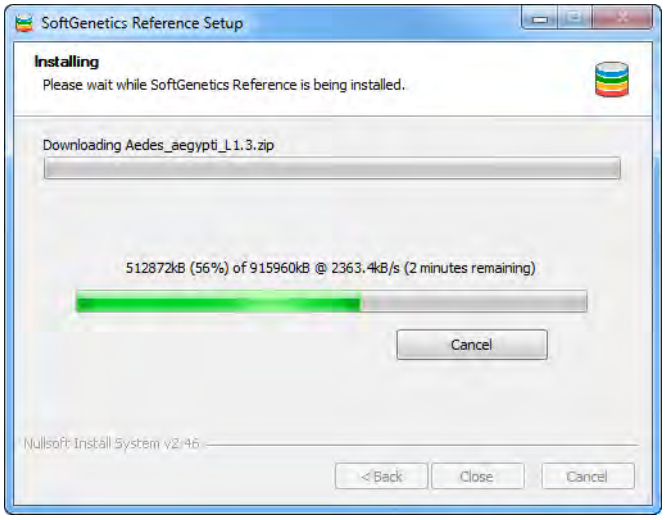
    - Make enough space in the installation directory, and then repeat Step 9.
    - Select fewer references to install, and then repeat Step 9.

11. Click Yes to begin the import.

    The Installing page opens. The page shows the status of downloading each indexed reference file. See Figure 2-24 on page 31.

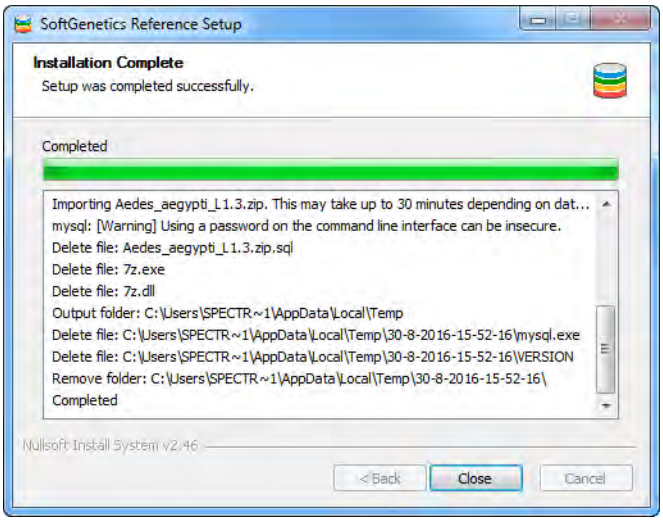
If you encounter any problems during the downloading and importing of the selected reference files, contact tech_support@softgenetics.com.

Figure 2-24: Reference Setup Wizard, Installing page for an in-progress installation



After all the selected preloaded references have been successfully downloaded and imported, the Installing page is updated with an Installation complete message.

Figure 2-25: Reference Setup Wizard, Installation page for a completed installation



12. Click Close.

    The preloaded references are now available for use.

The Build Preloaded Reference Tool (For NextGENe Only)

You use the Build Preloaded Reference tool to index any large reference sequence (≥ 250 Mbp) for any alignment application type, or any reference sequences that are to be used for the Transcriptome with Alternative Splicing Application type. You can use a BED file to create an index from an existing preloaded reference, or you can use any .fa, .fna, .fasta, GenBank, or pure sequence file to create the index.

Be aware of the following:

- *For Transcriptome analysis, you must use GenBank files so that annotation information can be included.*
- *If you need assistance in building your own index, or if you would like SoftGenetics to build an index for you, contact SoftGenetics directly.*

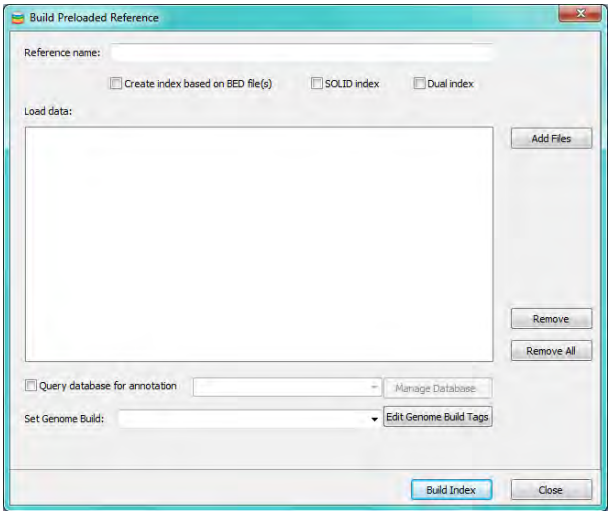
To use the Build Preloaded Reference tool with a BED file

You can use a BED file to recreate a part of the index for an existing whole genome file, for example, for exomes in a targeted region. You can use a BED file to recreate an index for any valid data type such as Illumina data, SOLiD data, and so on; however, if you use SOLiD data, you must explicitly indicate this.

1. On the Reference & Track Manager tool, click Build Reference.

   The Build Preloaded Reference dialog box opens.

Figure 2-26: Build Preloaded Reference dialog box



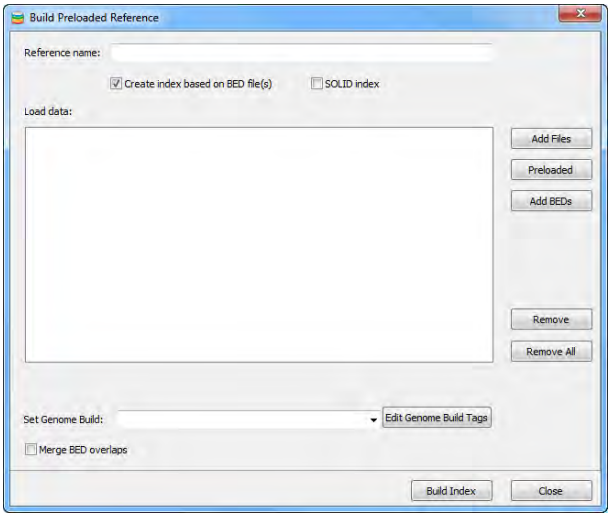
2. In the Reference name field, enter the name that is to be used for the reference.

The reference is saved to the Reference directory that is specified in the Storage Path settings from the Reference & Track Manager Tool.

3. Select Create index based on BED file(s).

   The Build Preloaded Reference dialog box is refreshed with options for creating an index using a BED file. A Merge BED overlaps option is also displayed.

Figure 2-27: Build Preloaded Reference dialog box BED file options



4. On the Genome Build dropdown list, select the genome build for the BED file. The newly created reference is saved under this genome build.
5. Optionally, click Edit Genome Build Tags to open the Genome Build Tags Editor dialog box, and edit any genome build tags as needed. See "To edit genome build tags" on page 14.
6. Do one of the following:
   - To avoid the overlapping of ROIs or amplicons from the loaded BED file, do not select Merge BED Overlaps.
   - To merge these ROIs or amplicons, select Merge BED overlaps.
7. If you are recreating an index using any data type other than SOLiD data, continue to Step 8; otherwise, select SOLiD Index, and then continue to Step 8.
8. In the Load Data pane, do the following:
   - Click Preloaded to select the reference that is to be recreated based on the BED file.
   - Click Add BEDs to browse to and select the BED files that are being used to recreate the index.

9. Click Build Index.

   The Output folder contains several output files, including the indexed reference file and an Excel CSV file, that detail the information about each contig reference position.

Figure 2-28: NextGENe Preloaded Reference tool output folder and files

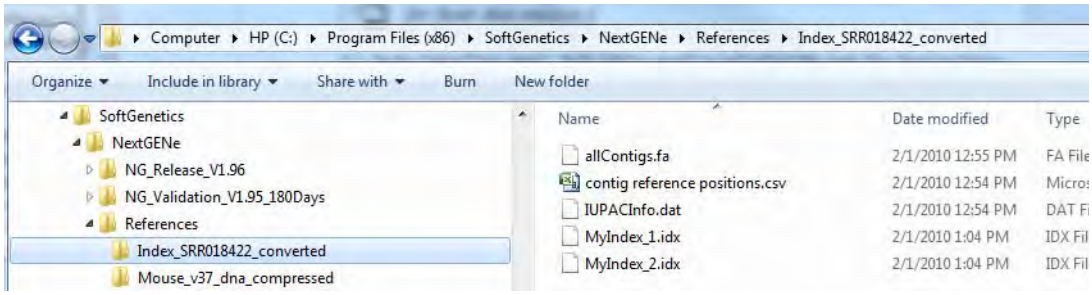


Figure 2-29: Sample contig reference position csv file

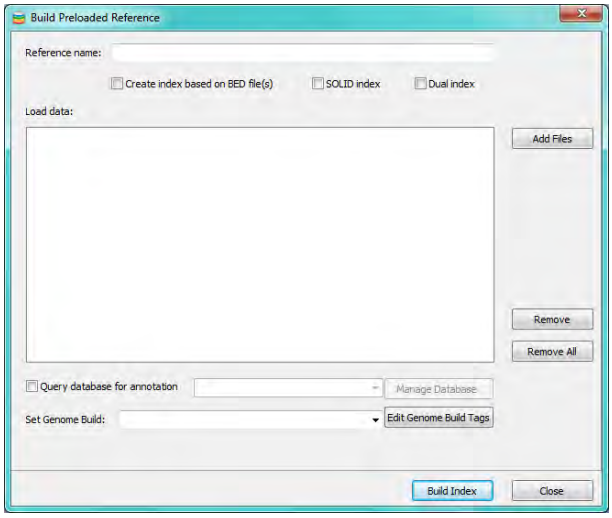
	A
1	ContigSizeChromChrom StartChrom EndReference Position
2	NT_SRR01842245chrFWGR3X101DYLGG0440
3	NT_SRR01842245chrFWGR3X101AEE3E458945

To use the Build Preloaded Reference tool to create a new index

1. On the Reference & Track Manager tool, click Build Reference.

   The Build Preloaded Reference dialog box opens.

Figure 2-30: Build Preloaded Reference dialog box



2. In the Reference name field, enter the name that is to be used for the reference.

The reference is saved to the Reference directory that is specified in the Storage Path settings from the Reference & Track Manager Tool. See

3. Do one or both of the following as appropriate:
   - To build an index to which you can align your SOLiD System data, select SOLiD Index.
   - To build two separate indices—a “standard” genome index and an index where the reference sequence is replaced with variant sequences based on variants reported in dbSNP—select Dual Index.

NextGENe can align sample files to both indices simultaneously, which can provide for faster data analysis.

4. In the Load Data pane, click Add Files to browse to and select the data files that are being indexed.
5. On the Genome Build dropdown list, select the genome build that is to be used for the new reference.
6. Optionally, click Edit Genome Build Tags to open the Genome Build Tags Editor dialog box, and edit any genome build tags as needed. See “To edit genome build tags” on page 14.
7. To include annotation information from an existing reference database, click Query database for annotation, and then select the appropriate database.

You can click Manage Database as needed to open the Process Options Settings dialog box and confirm or edit the MySQL settings. See Specifying Process Options in the NextGENe User’s Manual or Online Help.

8. Click Build Index.

   The Output folder contains several output files, including the indexed reference file and an Excel CSV file that detail the information about each contig reference position. See Figure 2-31 below and Figure 2-32 on page 36.

Figure 2-31: Build Preloaded Reference tool output folder and files

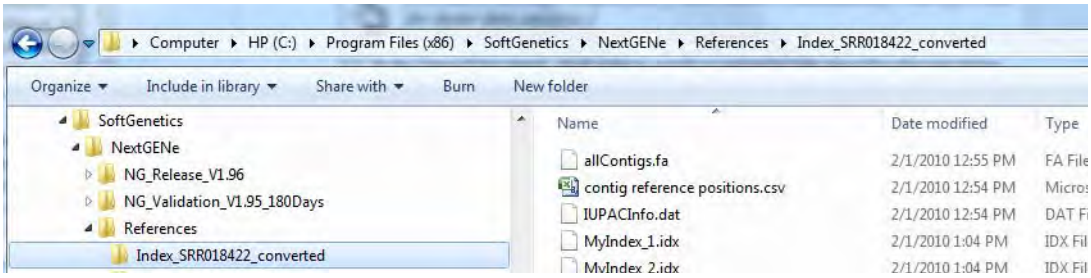
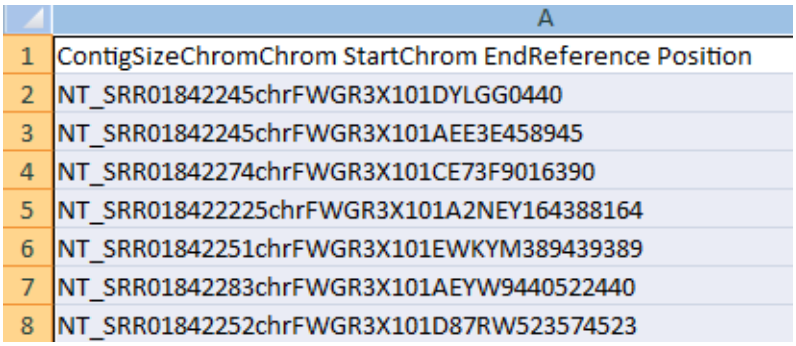


Figure 2-32: Sample contig reference position csv file



	A
1	ContigSizeChromChrom StartChrom EndReference Position
2	NT_SRR01842245chrFWGR3X101DYLGG0440
3	NT_SRR01842245chrFWGR3X101AEE3E458945
4	NT_SRR01842274chrFWGR3X101CE73F9016390
5	NT_SRR018422225chrFWGR3X101A2NEY164388164
6	NT_SRR01842251chrFWGR3X101EWKYM389439389
7	NT_SRR01842283chrFWGR3X101AEYW9440522440
8	NT_SRR01842252chrFWGR3X101D87RW523574523