Exon Array Computational Tool (ExACT)

Software User's Guide

Contents

List of Tables

List of Figures

Chapter I: Getting Started

1 Introduction

The Exon Array Computational Tool (ExACT) is a verified but not fully validated software tool for the analysis of exon array data. This tool takes, as the primary input, CEL files created in and extracted from GCOS and generates gene and exon level probeset results, such as signal estimates and detection p-values.

2 Overview of ExACT

Generally, use ExACT with exon arrays to:

- generate quality control report(s)
- normalize CEL file intensities
- generate gene-level signal estimates and detection p-values using meta probeset lists
- generate exon-level signal estimates and detection p-values
- merge signal estimates and/or detection p-values with various textual annotations including NetAffx CSV files
- generate expression graph files (egr) from probeset summary estimates for visualization of exon expression results in the Integrated Genome Browser (IGB)

ExACT is available as a pre-built Microsoft Windows application and is covered by this manual. ExACT source code can be compiled and run on Microsoft Windows, Mac OS-X, Solaris, and Linux. The source code release is available and supported via the Affymetrix Developer Network. The Windows application consists of several different command line applications and a GUI workflow application for executing these command line applications. The ExACT user can either run these command line programs directly, or ignore the command line programs and use the GUI workflow manager to perform exon array analysis. Mixing use of the command line applications and the GUI workflow manager on the same project files can be done, but is not supported. There is limited support through the Affymetrix Developer Network for use of the command line programs directly when the source code is released.

Because exon level analysis methods are still evolving, it is expected that the ExACT software will evolve along with the methods at a fairly rapid pace. While reasonable efforts will be made to maintain compatibility of newer ExACT software releases with previously generated ExACT results files, there may still be analysis methods, user workflows, and/or performance improvements which prevent backwards compatibility. To mitigate problems with backward compatibility and ease transitions during a relatively rapid release cycle, the ExACT software will allow the installation and use of multiple versions simultaneously.

> **NOTE**: ExACT consists of a collection of command line programs, documentation, and a GUI application to manage the workflow. Use of the term GUI Workflow or GUI Workflow Manager generally refers to the GUI application for doing exon array analysis. The GUI application consists of a main application window and a Workflow Wizard. The Workflow Wizard part of the GUI application contains a number of dialog boxes for setting analysis step parameters. Chapter II: and Chapter V: discus the main application window in detail and Chapter III: covers using the Workflow Wizard to control the analysis workflow.

3 Installation and Setup

This section covers the hardware and OS requirements, the software requirements, and the installation procedure for ExACT.

3.1 System Requirements

These are the recommended and preferred hardware and operating system requirements to run ExACT. Sufficient memory (RAM) to run ExACT is the most critical component, however fast CPU(s) and disks will significantly improve analysis run times.

ExACT will not use both processors if present, however, dual processor machines will allow the user to run multiple ExACT instances simultaneously. The hardware specifications below reflect specific configurations which are informally verified (but not validated). For example, machines with other disc configurations, Windows 2000, or Windows XP Service Pack (SP) 1 will probably run ExACT without problems.

	Recommended	**Preferred**
Processor	3.6 GHz Pentium 4	Dual Xeon 3.0 GHz
Memory (RAM)	2-4 GB	4 GB
Disk Drive Storage	250 GB 7200 RPM SATA	Dual 250 GB SATA in RAID1 configuration
Operating System	Windows XP Professional SP2	Windows XP Professional SP2 (32-bit)
Example System	Dell OptiPlex GX620 DT	Dell Precision 670 workstation

Table 1: System Requirements

In order to use ExACT, the user must also have the .Net 1.1 Framework installed. The .Net Framework can be installed using the Microsoft Update, which is available from the Windows **Start** button, and is typically listed under the **Programs** menu. For more information, see the Microsoft .NET website (http://msdn.microsoft.com/netframework/technologyinfo/default.aspx)

NOTE: The source code version of ExACT, available from the Affymetrix website under the Developer Network section is known to compile and run on other systems such as 32-bit and 64-bit versions of Linux on Intel and AMD hardware, Solaris on SPARC, and Mac OS-X on the Power PC. The source code release is provided with minimal support and is not covered by this User's Guide. Furthermore, the source code release on these platforms only supports a command line interface. The GUI workflow program only runs on Microsoft Windows systems.

CAUTION: The user should not have the ExACT software installed on the same workstation controlling the scanner or other instrumentation generating the data from GCOS or any raw data for this analysis.

3.2 Obtaining the ExACT Software

The ExACT software can be obtained from the Affymetrix website (http://www.affymetrix.com/products/software/specific/exact.affx) at no charge. The ExACT Windows release is covered by the Affymetrix free software license (see Appendix I: ExACT Software License, for more information).

NOTE: ExACT users should subscribe to the ExACT message service at http://www.affymetrix.com/products/software/specific/exact.affx to acquire the latest release information.

3.3 Installing ExACT

Once the installer (.msi extension) is downloaded, it can be used to install ExACT as outlined below.

NOTE: Multiple versions of ExACT can be installed simultaneously. This allows users to try out new releases of ExACT without having to uninstall previous versions. This also allows for smoother transitions when new releases do result in backward incompatibilities.

A. Start the installation process by double clicking on the installer icon with the mouse. The installer will display with the splash page as shown in Figure I-1, below.

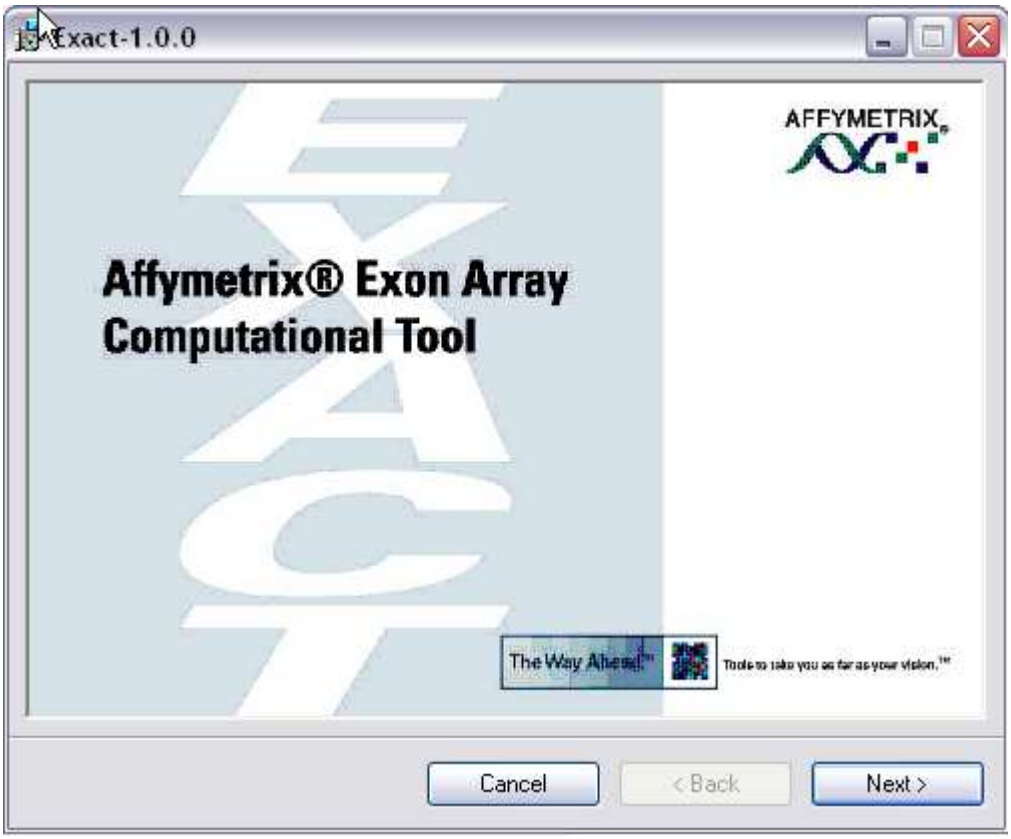

Figure I-1: ExACT Installer Splash Page

B. Click through the welcome screen, noting the information presented.

C. Read the ExACT Software License Agreement and either agree to the license by selecting the **I Agree** radio button, or cancel the installation by selecting **Cancel** as shown in Figure I-2. (For the license agreement text see

D. Appendix I: ExACT Software License on page 49 in this manual.)

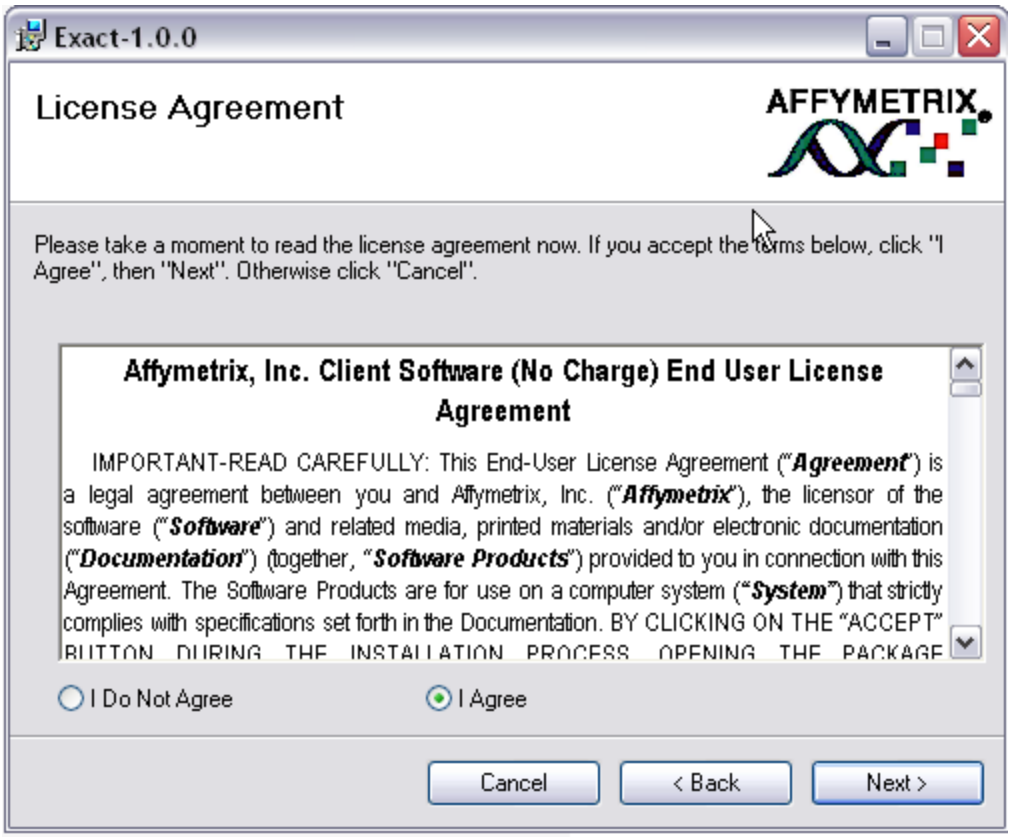

Figure I-2: ExACT License Acceptance Page

E. If the user accepts the license in the previous step and selects the **Next** button, an installation folder prompt will be displayed. In most cases the user should accept the default location as shown in Figure I-3, below.

F. By default, the programs will only be available within the account in which it was originally installed as directed by the **Just Me** radio button as shown in Figure I-3, below. To make the software available to other user accounts, select the **Everyone** radio button. Click the **Next** button to continue the installation.

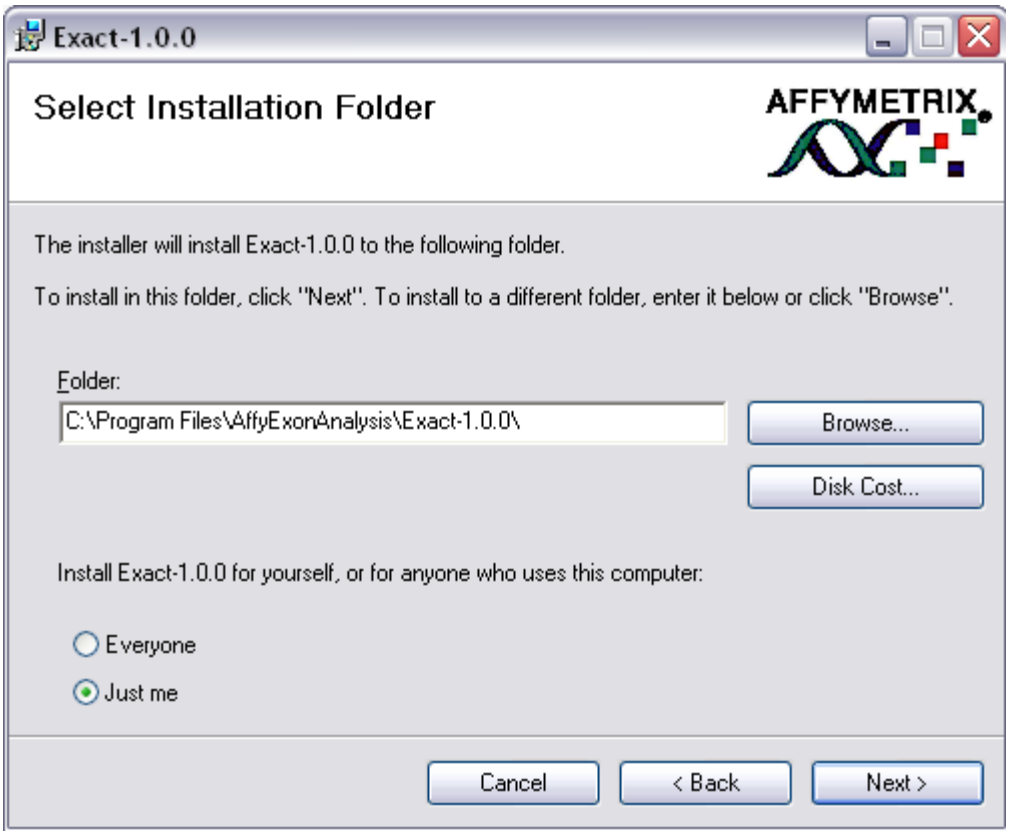

Figure I-3: ExACT Installation Folder Selection

> **NOTE**: The default installation folder includes the ExACT version. If the user has multiple versions of ExACT installed, they will all show up under the same AffyExonAnalysis item in the Programs menu accessible from the Windows Start button.

G. At this point, a window should be displayed indicating that the installer is ready to start installing files. Click the **Next** button and an installation progress window with an active progress bar will be displayed.

H. Once the installation is complete, the user should see the **Installation Complete** screen displayed as shown in Figure I-4, below. At this point the installation is complete and the user can click on the **Close** button to exit the installer.

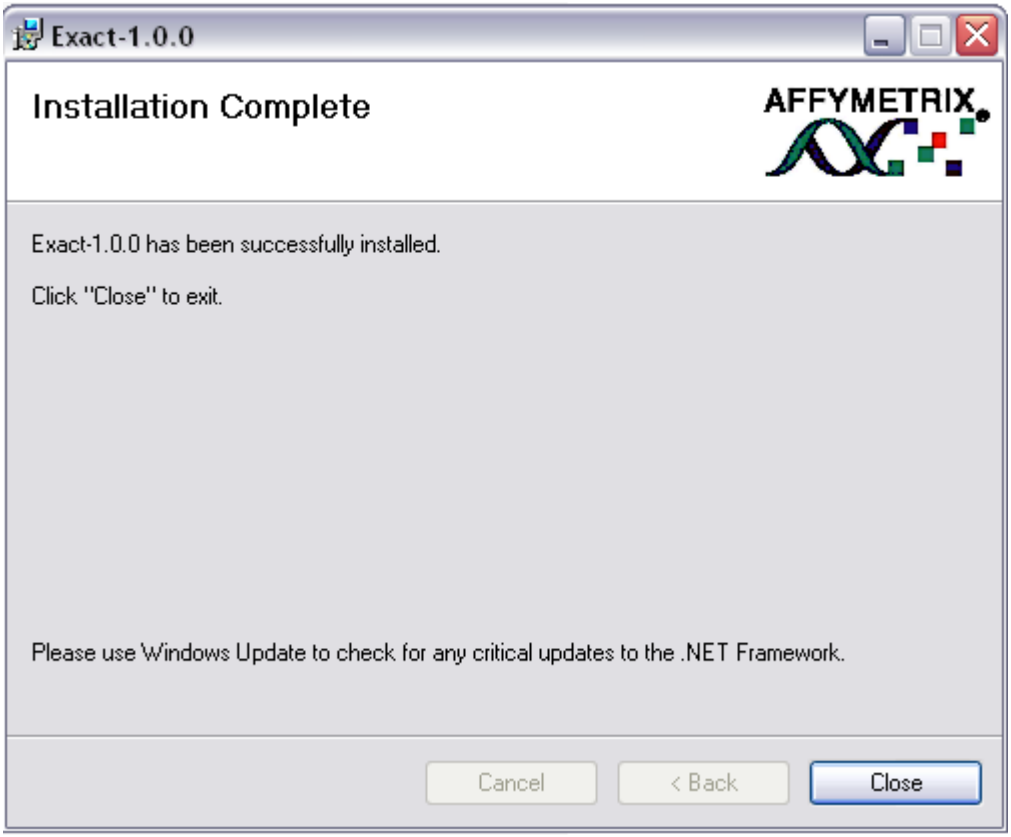

Figure I-4: ExACT Installation Complete

The installation process will install the software license, ReadMe file, and Change Log.

> **NOTE**: Make sure you review the ReadMe file and Change Log (available from the START menu and also from the Help menu within the ExACT GUI Workflow) for important notes about the software and changes (including potential backwards compatibility issues) specific to the release installed.

Documentation for each of the command line modules is installed at the same time as the ExACT application itself. Both the GUI Workflow Manager and associated ExACT command line documentation will be available from the **Start** menu after installation. In addition, the software will install the command line applications under the **Program Files** folder in a directory called **AffyExonAnalysis**, unless the user has indicated an alternative installation directory.

Chapter II: ExACT Workflow Wizard Main Window

1 Introduction

The ExACT GUI workflow is the supported interface for doing exon array analysis. The GUI workflow allows the user to create projects that specify a series of ExACT commands to analyze data from exon array experiments. This chapter explains the components and functions of the ExACT main window and introduces the ExACT workflow through the GUI.

To start the GUI, select the **Workflow Wizard** program under the **AffyExonAnalysis** item in the **Programs** folder from the **Start** menu.

2 ExACT Main Window

All ExACT analysis processing originates with the ExACT main window. The ExACT main window contains the **File** and **Help** dropdown menus, and the **Settings**, **Input Files**, **Output Files**, and **Results Log** tabs, each containing the respective contents of the tab subject in a page under that tab. There is also a truncated analysis tree view in the left window, however, it will remain empty until a new project is started or an old project is opened. The main window is shown with its component parts called out in Figure II-1, below.

> **NOTE**: Although most of the components of the ExACT main window will not be used in initializing the array analysis overview process, they will be useful once the analysis has been completed and the tabs are populated with information about the analysis.

Figure II-1: ExACT Main Window

2.1 Main Window Title Bar

The title bar of the main window is indicated in Figure II-1, above. The title contains the ExACT version number the user is working with at present.

> **NOTE**: Be sure to check the Affymetrix ExACT product website regarding product updates at (http://www.affymetrix.com/products/software/specific/exact.affx).

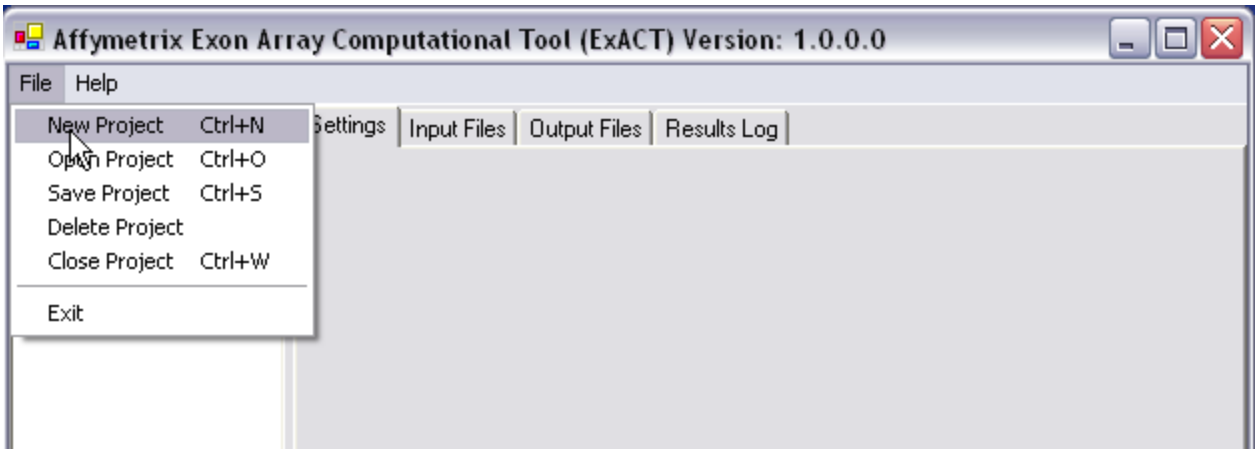

Figure II-2: File Dropdown Menu

2.2 Main Window File Menu

The **File** dropdown menu, shown in Figure II-2, contains a number of options for managing projects.

2.2.A Starting A New Project

This menu item allows the user to create a new analysis project. Selecting the **New Project** menu item will bring up the **Create New Data Analysis Project** dialog box (Figure III-2). See Chapter III: (Create A Data Analysis Workflow) for more information about how to create a new project.

2.2.B Opening A Project

This menu item allows the user to open an existing analysis project. When selected, browse to the appropriate project.xml file from a previous ExACT session. Selecting and opening this file will allow the user to view, add, and alter an analysis project.

2.2.C Saving A Project

This menu item will save the current state of the project.

> **NOTE**: A project is automatically saved when a project is closed or the user exits the ExACT Workflow GUI.

2.2.D Deleting A Project

This menu item allows the user to delete a project.

> **NOTE**: The project corresponds to a folder in the results directory with the same name (the project name is specified when creating a new project). The user can also delete the project and all its resulting output files by simply deleting this folder using the Windows Explorer. The user can also backup the project by simply copying this folder to an archive device such as a backup hard drive, or a DVD writer.

2.2.E Closing A Project

Select this menu item to save and close a project. The project can be opened at a later date in order to alter, remov, or add an analysis steps.

2.2.F Exiting ExACT

Select this menu option to save the currently opened project and to exit the ExACT Workflow GUI.

2.3 Main Window Help Menu

The **Help** dropdown menu can be used to view the online help and to get information about ExACT. The Help dropdown menu is shown in Figure II-3, below.

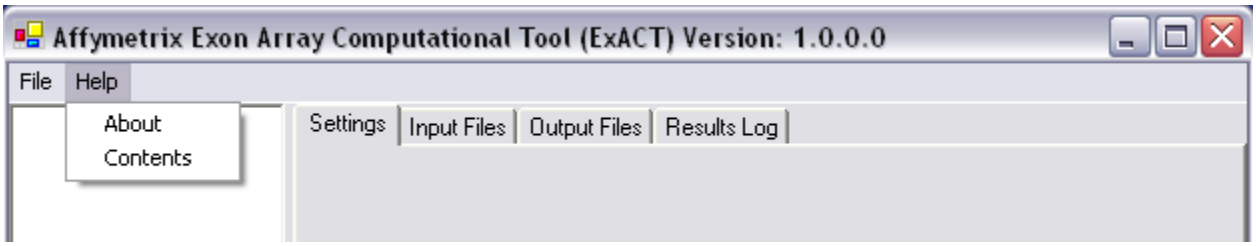

Figure II-3: Help Dropdown Menu

2.3.A About ExACT

Use the **About ExACT** message to check the version number of the software being used.

2.3.B Help Contents Page

Online help for ExACT is available via this menu option.

3 Project Information Tabs

The project information tabs (**Settings**, **Input Files**, **Output Files**, and **Results Log**) are used to display information about the currently selected analysis step within a project. These tabs are empty when there is no active project. See Chapter V: 3 for more information.

4 Project Tree View

Once a project has been created or executed, or a previously run project has been opened, the user can use the Main Window for a variety of tasks, including:

- View the settings used to run the analysis
- Review the report log for each analysis step
- Open results folders
- Create new steps
- Clone existing steps
- Delete steps

See Chapter V: 2 , (Project Tree View), for more information

Chapter III: Create A Data Analysis Workflow

1 Introduction

This chapter explains how to setup a new probe level analysis using the ExACT Workflow Wizard. The high-level workflow consists of:

- Defining a new project
- Importing CEL files
- Normalizing CEL files
- Generating probe summaries (i.e., PLIER signal estimates and p-values)
- Adding textual annotations to the summaries
- Generating visualization files from the summaries for the Integrated Genome Browser (IGB)

This chapter discusses the specifics of each step as well as how to use the online help associated with the more complex steps.

2 Online Help for Wizard Steps

Each of the wizard windows starting with the Normalization step is related to a particular command line application which the GUI Workflow Wizard manager will execute after the **Finish** button is clicked. Each of these command line modules has online documentation which can be accessed using the question mark button on the upper right corner of each dialog box.

Most of the entry boxes and buttons on the Wizard dialog boxes correspond to a particular command line option. To view the command line option associated with a particular text box or button on the GUI wizard, move the pointer over the labeling next to the field on a dialog box. The command line parameter corresponding to the function labeled will be displayed immediately underneath the labeling. An example of this is shown in Figure III-1, below, where the Summary Method on the wizard GUI is shown to correspond to the **–qmethod** command line option.

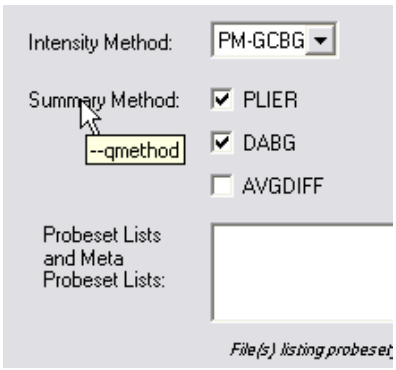

Figure III-1 Command Line Parameter Pop-up Hints

The combination of the online command line documentation (via the question mark button) and the mouse over command line parameter allows the user to navigate to more detailed information regarding a particular option or parameter.

3 Create New Data Analysis Project

ExACT exon array analysis is organized by projects. The first step in analyzing exon array CEL files is to create a new project by defining a storage location and a project name. An optional description may be provided. Once a project is defined all subsequent analysis results will be stored in files located in the project **Results Folder**. The project results folder is a directory with the same name as the project name located in the storage location specified during this step. See Chapter V: 3.4 (Results Log Tab), for more information.

To begin a new data analysis project, select the **New Project** menu item from the **File** dropdown menu on the **ExACT** main window. The **Create New Data Analysis Project** dialog box will be displayed as shown in Figure III-2, below.

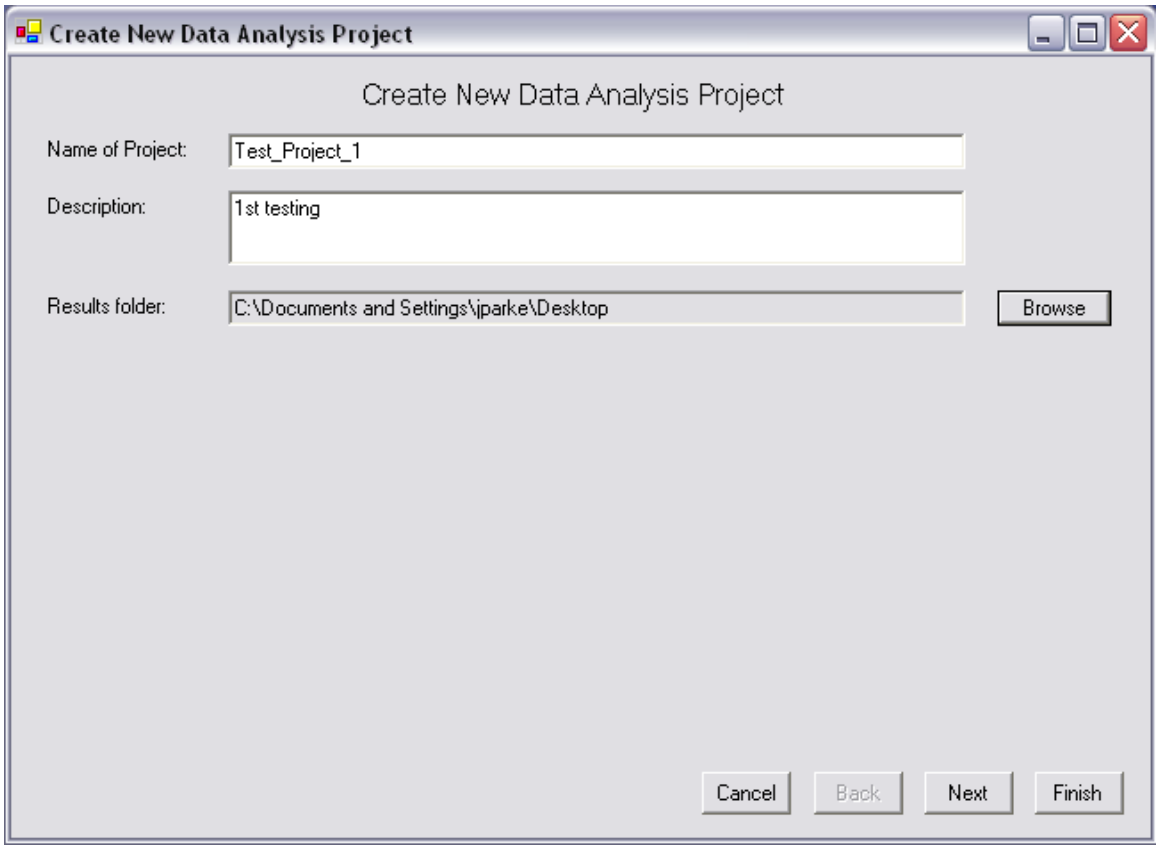

Figure III-2: Wizard Step 1 - Create New Data Analysis Project

3.1 Naming the Project

> **NOTE**: The name of the project is used as the directory name under the Results Folder. All the results generated by ExACT and associated with this project will be stored in this directory.

A. Direct the pointer to the **Name of Project** field.

B. Type the name of the project in the field. The project name is displayed as shown above.

> **NOTE**: Underscores will be displayed in the Name of Project field if the space bar is typed as shown in Figure III-1, above.

3.2 Describing the Project (Optional)

A. Direct the pointer to the **Project Description** field.

B. Type the description for the project in the field. The project description is displayed as shown in Figure III-2, above.

> **NOTE**: The description is not used by the analysis software, but can be useful to the researcher as a place to indicate the purpose of this particular project or special comments regarding the samples being analyzed.

3.3 Determining the Project Folder Destination

> **NOTE**: The results for the new project will be stored in a subfolder corresponding to the project name. Thus one can use the same Results Folder for multiple projects.
>
> **NOTE**: The Results Folder should be located on a local disk or fast network storage for optimal performance.
>
> **NOTE**: A project can be backed up by simply copying the project folder under the Results Folder to an archive device such as a backup external hard disk.

A. Use the **Browse For Folder** dialog box to select the **Results Folder** where data derived from this data analysis project will be contained.

> **NOTE**: For optimal performance, the user should select a folder located on a local disk rather than a network location.

B. Click the **OK** button after the **Results Folder** destination has been selected. The path to the destination for the **Results Folder** selected will be displayed as shown in Figure III-2, above. If no destination is selected, Windows will default to the **My Computer** folder.

C. To cancel the destination selection process and return to the **Create New Data Analysis** dialog box, click the **Cancel** button on the **Browse for Folder** dialog box.

3.4 Proceeding to the Next Step

D. To proceed to the next step in the ExACT workflow, selecting .CEL files to be analyzed, click the **Next** button on the **Create New Data Analysis** dialog box. The **Import CEL Files** dialog box will be displayed.

E. To cancel the creation of a new project and return to the main window, select the **Cancel** button.

4 Import CEL Files

The **Import CEL Files** dialog box allows the ExACT user to select which CEL files to analyze. The **Import CEL Files** dialog box will be displayed after the appropriate fields in the **Create New Data Analysis Project** dialog box have been filled, and the **Next** button has been clicked. Both text and binary CEL files can be imported, however, binary CEL files are preferred for improved performance and reduced memory use. CEL files must be extracted from GCOS and transferred to the ExACT workstation prior to running this step. The GCOS File Copy tool can be used to extract binary (XDA) CEL files from GCOS for use by ExACT. When released, the Data Transfer Tool (DTT) version 1.1 will replace the GCOS File Copy Tool.

The "raw" CEL files from GCOS can be analyzed by ExACT in their current location (imported by reference) or by copying them to the Project Folder. To conserve disk space, it is recommended that CEL files be imported by reference (default). If the user wants to preserve the raw input CEL files in addition to the rest of the project results, then the user should use the **import by copy** option.

The **Import CEL Files** dialog box contains fields for naming the step, describing the step, buttons to add or remove CEL files, a checkbox to copy the files selected here into the project folder, an active window where the CEL files selected can be viewed, and the fields where the CEL files selected can be grouped according to replicates.

CEL files can be organized into replicate sub-groups using the **Group Name** field. Currently, the grouping information is only used by the normalization if the **Normalize by Group** radio button is selected. The user may want to group technical and/or biological replicates into the same group by associating them with an identical group name.

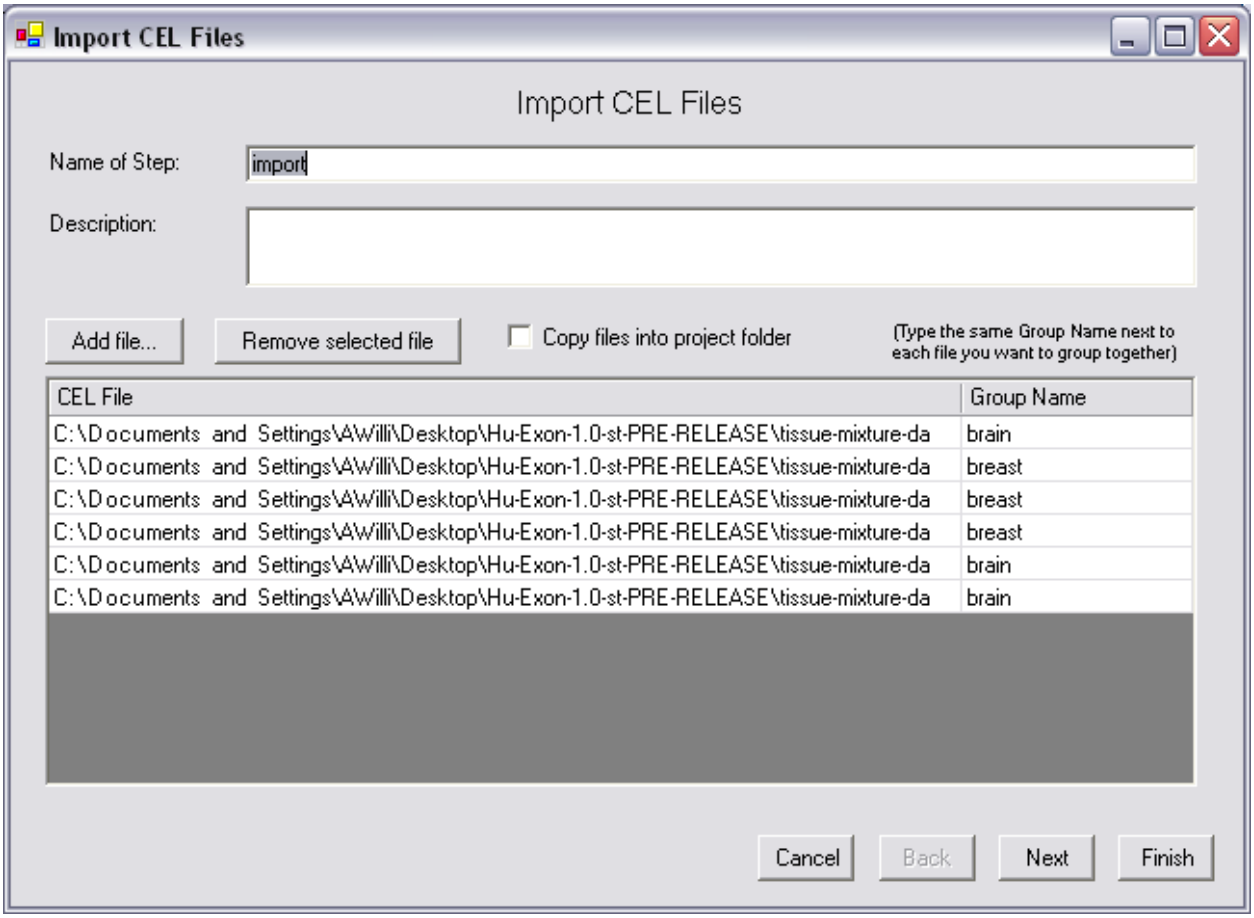

Figure III-3 Wizard Step 2 – Import CEL Files

4.1 Naming the Step

> **CAUTION**: The step name determines the folder name on the disk. Avoid the use of special characters. Generally, ExACT will not allow the use of these characters in the step name.
>
> **NOTE**-: No two steps within the same project can have the same name.

A. Type the name of the Import CEL files into the **Name of Step** field. Note that the word **import** is displayed first in the **Name of Step** field as a default, and can be edited by the user for a more detailed step name as required. The name typed is displayed as shown in Figure III-3, above.

B. Note that underscores are automatically entered in the **Name of Step** field when the space bar is keyed.

4.2 Describing the Step (Optional)

A. Type the description of the step the user is performing into the **Description** field (Figure III-3).

4.3 Adding Files

A. Click the **Add file** button. The default folder or location for files to be selected is displayed.

B. Select the files to be processed with ExACT by clicking on them. Multiple files can be selected by holding down the **Shift** and/or **CTRL** key. The files selected will be displayed in the active window of the **Import CEL Files** dialog box as shown in Figure III-3, above. To remove a file that has been selected, select it in the active window by clicking on it, then click the **Remove file** button. The file will be deleted from the list.

> **NOTE**: In addition to importing raw CEL files from GCOS, the user can also import normalized CEL files from a previous ExACT analysis.

4.4 Copying CEL Files Into the Project Folder:

A. To save a copy of the CEL files to be analyzed by ExACT in the project folder, click the **Copy files into project folder** checkbox on the **Import CEL Files** dialog box.

> NOTE: The destination for the Project Folder is the same folder selected and named in Chapter III: 3.3 above.

B. The files on the list will be copied into the pre-selected folder for this project once the **Finish** button is clicked. In most cases users will want to define additional steps before clicking the **Finish** button.

4.5 Naming Files in the Group Names (Optional)

A. Using the pointer, select the group name field to the right of the file with its group membership to be changed. This will select the current value.

B. Type in the new group name for this file. Multiple files that are to be grouped together should have the same group name.

C. The group names will be displayed as shown in Figure III-3.

4.6 Canceling This Step

A. Click the **Cancel** button on the dialog box to cancel this step.

B. The user is returned to the ExACT main window without saving any analysis step modifications to the project.

4.7 Returning to the Previous Step

A. Click the **Back** button on the dialog box to return to the previous step.

B. The previous step in the ExACT process is displayed.

4.8 Proceeding to the Next Step

A. To proceed to the next step in the ExACT analysis project, click the **Next** button.

B. The **Define next step** dialog box with the completed steps displaying a check mark to the right will be displayed as shown in Figure III-4, below. Note that the step that follows this procedure in the ExACT process, **Normalization**, has a radio button selected. To do probe summarization (i.e., calculate PLIER signal) using the un-normalized CEL files as they were imported, select the **Probe Summarize** option rather than the default **Normalization** option.

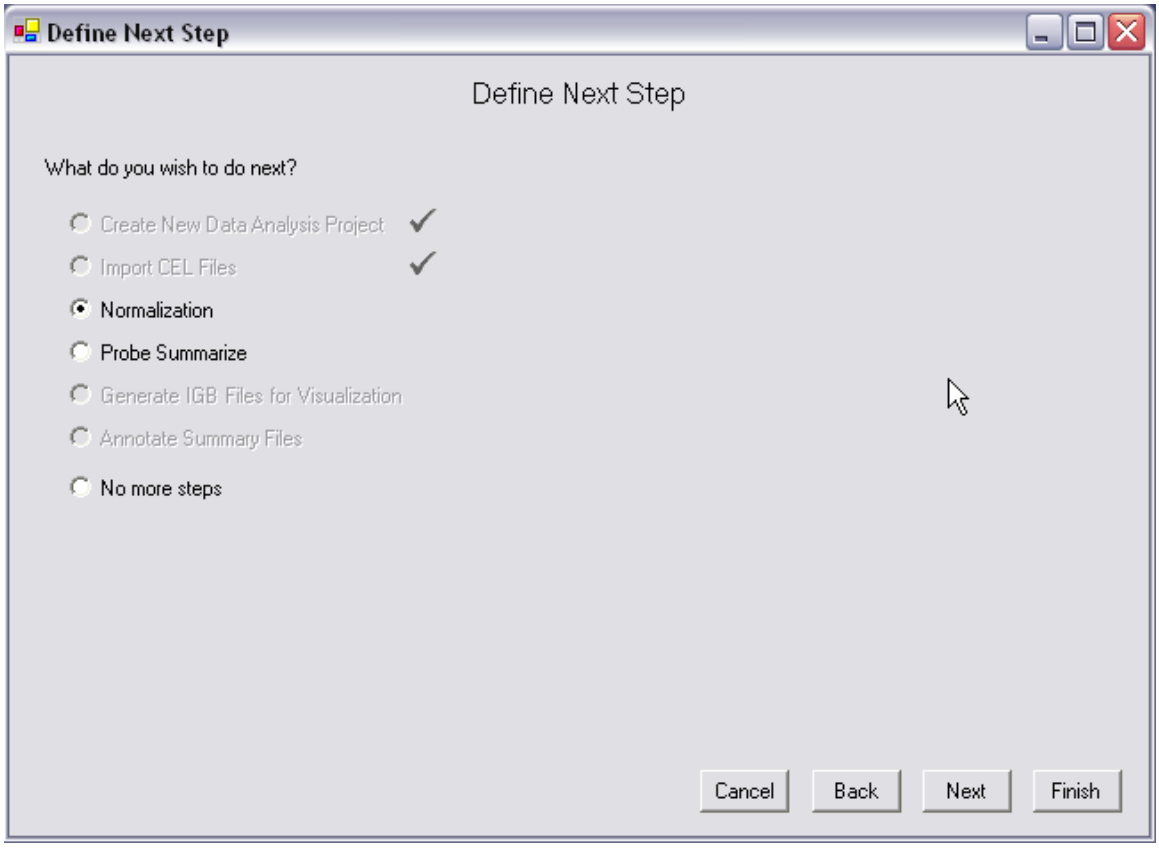

Figure III-4 Define Next Step After Import

5 Normalization

The normalization step supports several different normalization methods and allows the user to specify the CEL file output format. The normalization step takes a set of CEL files (either the ones exported from GCOS or CEL files from a previous normalization step), applies a normalization transformation, and then writes out new CEL files.

> **NOTE**: The normalization step can be skipped by selecting probe summarize as the step to follow CEL file import within the next step dialog box. Also note that multiple normalization methods can be linked together by selecting the normalization step again once the Next button is clicked.

The **Normalization** dialog box is shown below.

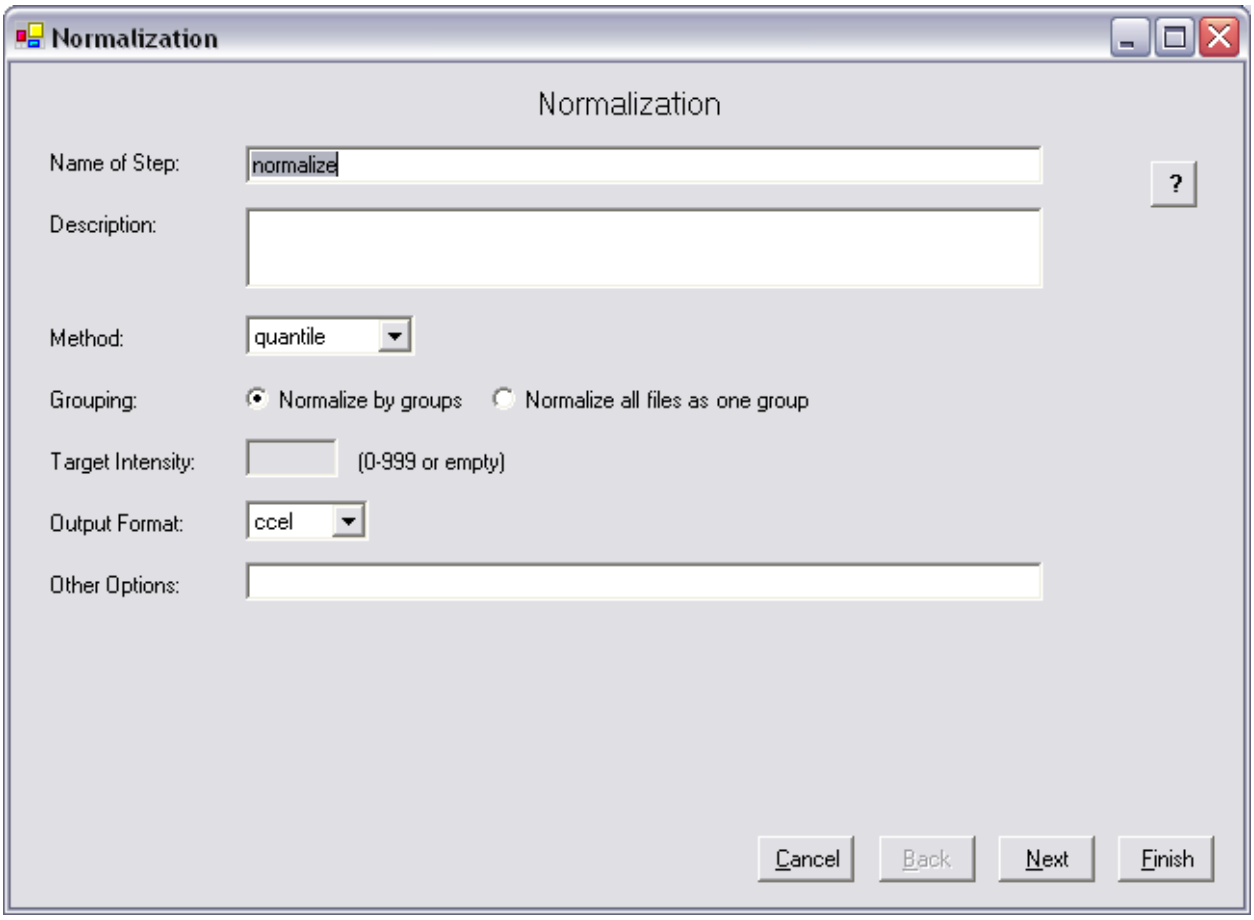

Figure III-5: Wizard Step 3 – Normalization

5.1 Naming the Step

A. Type the name of the step into the **Name of Step** field. Note that the word **normalize** is displayed first in the **Name of Step** field as a default. The name typed is displayed as shown in Figure III-5, above. This field can be edited by the user for a more detailed step name as required.

B. Note that underscores are automatically entered in the **Name of Step** field when the space bar is keyed.

5.2 Describing the Step (Optional)

A. Type the description of the step the user is performing into the **Description** field. The description typed is displayed in the **Description** field as shown in Figure III-5, above.

B. Note that spaces, and no underscores, are automatically entered in the **Description** field when the space bar is keyed. The only character count limitation is the size of the field, so very large and detailed descriptions can be entered.

5.3 Normalization Method

Several different normalization methods are supported. The default is **sketch** normalization which does a quantile-like normalization but in a memory-efficient manner allowing for the normalization of a large number of CEL files. Sketch normalization essentially uses an approximation of the intensity distribution rather than the full intensity distribution to do a quantile-like normalization.

The full **quantile** normalization method is also implemented, but only a limited number of CEL files can be normalized due to the memory requirements associated with this implementation. Roughly 10 CEL files can be normalized with the quantile method by ExACT.

There is also a **median** linear scaling method that adjusts the median intensity measured on each array to equal a pre-defined target intensity (or the median of CEL median intensities if no target is specified).

A. To select the type of normalization for this project, direct the pointer to the **Method** dropdown menu on the **Normalization** dialog box.

B. Click on the normalization dropdown menu down pointing arrow and select the normalization method of choice.

5.4 Grouping

The normalization method selected can be applied to all the CEL files imported for this analysis by default, or they can be applied to each group, as specified during the import CEL file step separately. For example, the user might want to quantile-normalize replicates within each group together, followed by a second normalization step with the entire quantile-normalized CEL files in that group median-normalized over the entire data set.

NOTE: Grouping can be used with any of the normalization methods.

5.5 Target Intensity Options (Optional)

The **Target Intensity** field allows the ExACT user to select the specific signal intensity target for median normalization. To select a specific target intensity, enter a numeric value greater than 0 and less than 999 in the **Target Intensity** field per the text instructions to the right of the field. If the median method is selected and no target is selected, then the median of the chip intensity medians is used as the target.

NOTE: Target Intensity only pertains to the Median normalization method.

5.6 Output Format Options

The normalization step allows users to specify the output CEL file format. In most cases the user should use the default **ccel** format which minimizes the file size and also reduces the memory requirements for subsequent steps. The formats are as follows:

- **ccel:** An ExACT internal format which only stores the feature intensity. The number of pixels, standard deviation of pixels, etc… are not retained. The feature intensity is stored as an integer (unsigned short) to further reduce the memory requirements.
- **xcel:** This is the standard XDA binary CEL file format. This format retains information not used within ExACT and also stores the intensity as a floating point number (float).
- **tcel:** This is the standard text format. Use of this format is discouraged because of the increased file size and the fact that these text CEL files cannot be used efficiently in the downstream methods.

A. To select the output format for this project, direct the pointer to the **Output Format** dropdown menu on the **Normalization** dialog box.

B. Click on the dropdown menu down pointing arrow and select the output format of choice.

5.7 Other Options

Enter any other options as notes to the ExACT user in the **Other Options** field at the bottom of the **Normalization** dialog box. Additional options that the user might want to consider are listed in the command line documentation available via the question mark button (see Chapter II: 2.3)

> **NOTE**: The **Other Options** field is essentially for advanced options. Most users can ignore this field.

5.8 Canceling This Step

A. Click the **Cancel** button on the dialog box to cancel this step.

B. The user is returned to the ExACT main window without saving any analysis step modifications to the project.

5.9 Returning to the Previous Step

A. Click the **Back** button on the dialog box to return to the previous step.

B. The previous step in the ExACT process is displayed.

5.10 Proceeding to the Next Step

A. To proceed to the next step in normalization process, click the **Next** button.

B. The **Define next step** dialog box with the completed steps displaying a check mark to the right will be displayed as shown in Figure III-6, below. Note that the step that follows this procedure in the ExACT process, **Probe Summarize**, has a radio button selected. The user can also perform second or additional rounds of normalization by selecting the **Normalization** option again.

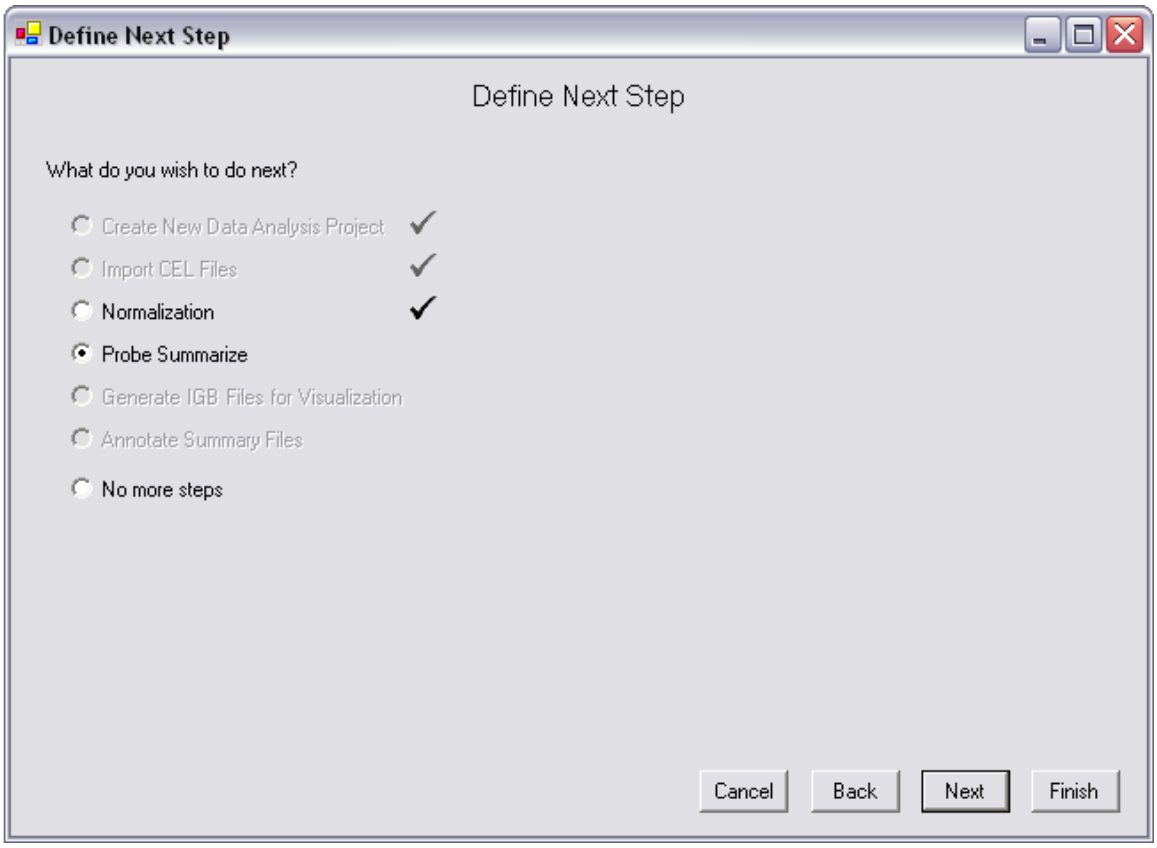

Figure III-6 Define Next Step After Normalization

6 Probe Summarize

The summarization step allows the user to define how various probeset summaries are done. For example, this is the step where exon or gene-level signal estimates are computed. This is also the step where a report file is generated containing metrics useful for assessing the data quality and identifying outlying experiments.

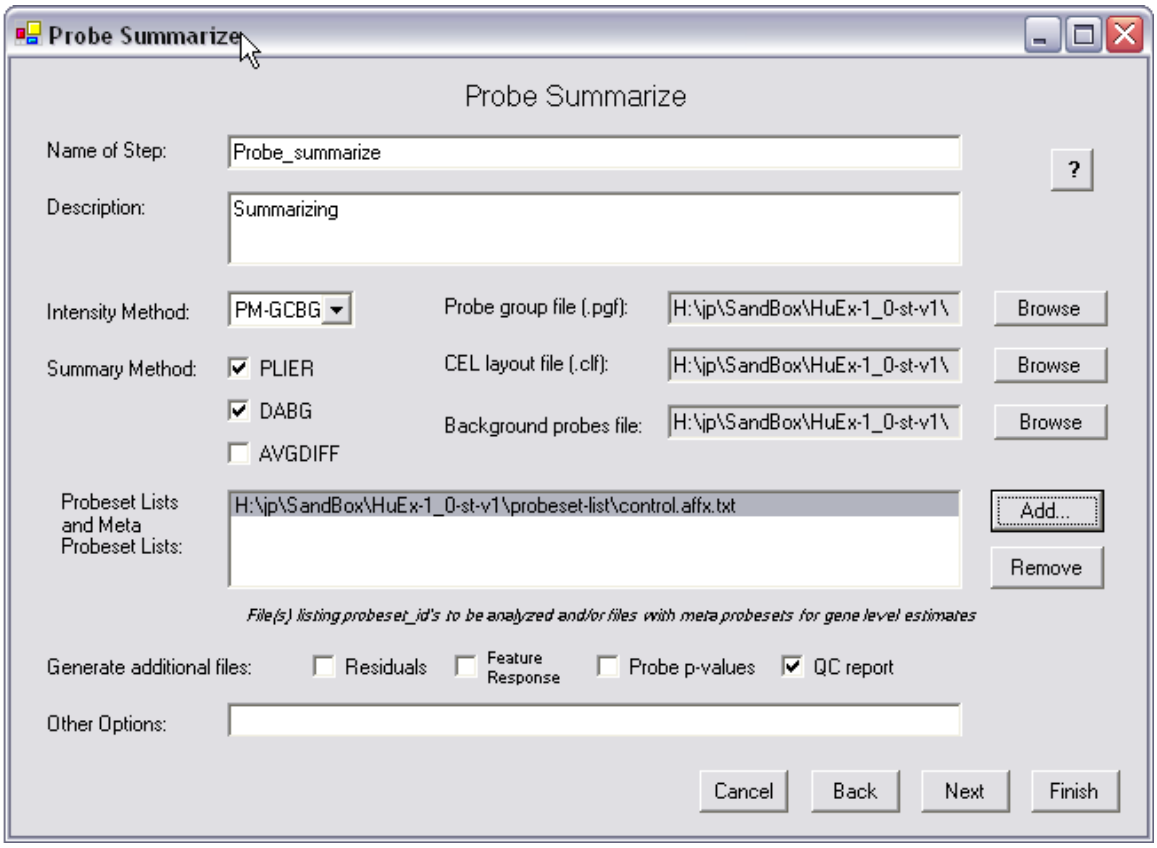

Figure III-7 Wizard Step 4 – Probe Summarize

6.1 Naming the Step

A. Type the name of the step into the **Name of Step** field. Note that the word **summarize** is displayed in the **Name of Step** field as a default. The name typed is displayed as shown in Figure III-7, above. This field can be edited by the user for a more detailed step name as required.

B. Note that underscores are automatically entered in the **Name of Step** field when the space bar is keyed.

6.2 Describing the Step (Optional)

A. Type the description of the step the user is performing into the **Description** field. The description typed is displayed in the **Description** field.

B. Note that spaces, and no underscores, are automatically entered in the **Description** field when the space bar is keyed. The only character count limitation is the size of the field, so very large and detailed descriptions can be entered.

6.3 Intensity Method

The intensity method is how the user configures the background correction. The default method is to use the background probes to do a GC composition-based background correction (PM-GCBG). The options are as follows:

- **PM**: This enables a PM-only type analysis of probesets. The background and mismatch probes are ignored. No background correction is performed.
- **MM**: This enables an MM-only analysis which might be appropriate for a study focused on highly expressed genes. No background correction is performed and the perfect match probes are ignored. This is not applicable to exon arrays which do not contain specific mismatch probes. An error message will be generated in the Results Log for probesets lacking specific mismatches.
- **PM-MM**: This enables the use of PM probes as measures of the target and the specific mismatch probes (MMs) as measures of probe-specific background. This is not applicable to exon arrays which do not contain specific mismatch probes. An error message will be generated in the Results Log for probesets lacking specific mismatches.
- **PM+MM**: This method treats both the PM and MM probes as measures of target. No background correction is performed. This is not applicable to exon arrays which do not contain specific mismatch probes. An error message will be generated in the Results Log for probesets lacking specific mismatches.
- **PM-GCBG**:(default, this is **PM** minus **GCBG**) This method uses a set of background probes (specified using the background probe file) to perform background correction. The background probes are also used as the distribution for computing the DABG values.

> **NOTE**: Only the default intensity method, PM-GCBG, allows all three summary methods available to be selected either individually or collectivly selection. All four other background correction choices will disable the DABG summary method, and the Probe Summarize dialog box will reflect this with the DABG function disabled as indicated by the DABG checkbox displayed as grayed-out.
>
> **NOTE**: The PM-MM, MM, and PM+MM methods are only valid for chip designs which contain specific mismatch probes. For example, the Human Exon 1.0 ST array does not, and as a result only the PM or PM-GCBG methods are valid.

A. To select the intensity method for this project, direct the pointer to the **Intensity Method** dropdown menu on the **Probe Summarize** dialog box as shown in Figure III-7, above.

B. Click on the dropdown menu down pointing arrow and select the intensity method for this project from the dropdown menu.

6.4 Summary Method

The summary methods define how probe level intensities are condensed into a single measure.

A. To select the summary method for this project, direct the pointer to **Summary Method** area of the **Probe Summarize** dialog.

B. Click the checkbox of the **Summary Method**:

- **PLIER** (Probe Logarithmic Intensity Error), generates a signal estimate for the probeset using a model fit as described in the PLIER Technical Note.
- **DABG** (Detected Above BackGround), generates a detection p-value based on comparing the perfect match probe intensity to the intensity distribution provided by background probes sharing the same GC content as the PM probe under consideration. This method is only available with the PM-GCBG method.
- **AVGDIFF** (Average Difference), this method computes a simple average difference between the target measures, such as PM probes, and the background measures. (For the Human Exon 1.0 ST Array, the median of the matched background probes for a given perfect match probe is used when the PM-GCBG intensity method is selected.) In most cases the user will

probably want to use **PLIER** signal estimates rather than the simple **AVGDIFF** signal estimates, as **PLIER** provides statistically a more robust estimate of the probe set level analysis than **AVGDIFF**.

> **NOTE**: In most cases the user will probably want to use **PLIER** signal estimates rather than the simple **AVGDIFF** signal estimates, as **PLIER** provides a statistically more robust estimate of the probe set level analysis than **AVGDIFF**.
>
> **NOTE**: Technical details regarding various analysis methods and algorithms are available via the Exon Array Analysis page on the Affymetrix website.

C. Select the summary method for this project. A check mark will be displayed in the checkbox(es) selected.

> **NOTE**: Multiple summary methods can be made simultaneously. A separate text result file will be generated for each summary method selected.

6.5 File Selections

> **NOTE**: Default .pgf, .clf, and .bgp files are provided by Affymetrix with each array. These can be downloaded from the Affymetrix website. Do not make any changes to these files for basic analysis.

A. A probe group file (pgf)), CEL layout file (clf)), and background probes file (bgp) should be selected by clicking on the **Browse** button to the right of each of the **probe group file (.pgf**), **CEL layout file (.clf),** and/or **Background probes file (.bgp)** fields.

> **NOTE**: The bgp file is only required when using the **PM-GCBG** intensity method.

B. When the **Browse** button is clicked, the **Open browser** dialog box with the directory hierarchy is displayed. Choose the file to be summarized for the user's project from the folder. The path to the selected file is displayed in the field.

6.6 Probeset Lists and Meta Probeset Lists Selection

This option serves two different functions dependent on the types of list added to the box.

- By adding **probeset lists** to this box, the probe summarization step will perform exon-level analysis on only the probesets defined in those files. Therefore, the user can effectively customize and filter the analysis to focus on a subset of the array content. Default probeset lists are provided by Affymetrix. Probeset lists can also be generated by querying the array content on the NetAffx Analysis Center.
- By adding **meta probeset lists**, gene-level signal estimates can be generated. The meta probeset list defines what probesets belong to a ***gene***. By default, three meta probeset lists can be downloaded from the Affymetrix website to support the gene-level analysis: core meta probesets, extended meta probesets and full meta probesets. These meta probeset lists can be manually modified to change the definition of each gene by users.

The user can import a mixture of probeset lists and meta probeset lists. However, care should be taken when adding multiple meta probeset lists. Specifically, the user should not mix meta probeset lists that define the same meta probeset ID for different collections of exons. As an example, the user should not add both the core meta probeset list and the extended meta probeset list, as both of these lists use the default transcript cluster ID as the meta probeset ID.

> **NOTE**: Add a meta probeset list to this field to perform gene-level signal estimates.
>
> **CAUTION**: If no probeset list or meta probeset lists are selected, ExACT will generate probe summaries for the entire chip during the execution phase. This will require approximately 40 minutes per CEL file for the Human Exon 1.0 ST arrays. It is recommended that an initial analysis focused on just control probesets using the control probeset lists prior to analyzing the entire chip be performed in order to quickly identify outlier arrays.

A. Probeset lists and meta probeset lists can be selected for probe summarization by clicking the **Add** button to the right of the field.

B. A windows browser directory hierarchy is displayed when the **Add** button is clicked. Select the appropriate probesets or meta probesets from the directories. Multiple files can be selected in the browser directory by using the Shift and/or Ctrl keys. The path to the folder or file(s) selected is displayed in the field.

6.7 Generating Additional Files

A. By default, a QC report, the DABG files, and the PLIER summary files are created during summarization. To generate additional summarization files from the file types indicated at the bottom of the **Probe Summary** dialog box for this project, click the checkbox to the left of the file type labeled.

B. Click the checkbox for information the user wants written to a file:

> **NOTE**: For each quantification method (ie PLIER,DABG,…) selected a summarization file is automatically created (ie plier.summary.txt, dabg.summary.txt, …).

- **Residuals:** Describe the PLIER signal estimate fit to the PLIER model.
- **Feature Response:** Describes the probe affinity corrections used for the PLIER model fit
- **Probe p-values:** from the DABG method
- **QC Report:** generates a summary report of the data set which is useful for basic QC and identifying outliers (this option is selected by default).

> **NOTE**: Many of the QC report values calculated will depend on the probe lists defined in the Probe Lists and Meta Probe Lists fields. These QC metrics will not be comparable across projects unless the QC report was generated from the same probeset lists. More information on the use of these files can be found on the Exon Array Analysis page on the Affymetrix website.

6.8 Other Options

Enter any other options as notes to the ExACT user in the **Other Options** field at the bottom of the **Probe Summary** dialog box. Other options that the user might want to consider are listed in the command line documentation available via the question mark button.

> **NOTE**: The **Other Options** field is essentially for advanced options. Most users can ignore this field.

6.9 Canceling This Step

A. Click the **Cancel** button on the dialog box to cancel this step.

B. The user is returned to the ExACT main window without saving any analysis step modifications to the project.

6.10 Returning to the Previous Step

> **NOTE**: Unless all files have been selected for the appropriate fields in the three file selection parts of Probe Summarization, ExACT will prompt the user to enter those files omitted for the appropriate before returning to the previous through the Back button. The only exception to this is if an intensity method other than the default, PM-GCBG, has been selected, which disallow the selection of a background probe file (bgp) for analysis

A. Click the **Back** button on the dialog box to return to the previous step.

B. The previous step in the ExACT process is displayed.

6.11 12. Proceeding to the Next Step

A. To proceed to the next step in the ExACT process, click the **Next** button.

B. The **Define next step** dialog box (Figure III-8) with the completed steps displaying a check mark to the right will be displayed. Note that the step that follows this procedure in the ExACT process, **Generate IGB files for Visualization**, has a radio button selected. The user can skip the generation of visualization files by selecting to annotate the summary files, or by clicking the **Finish** button, which will start the execution of the various analysis steps.

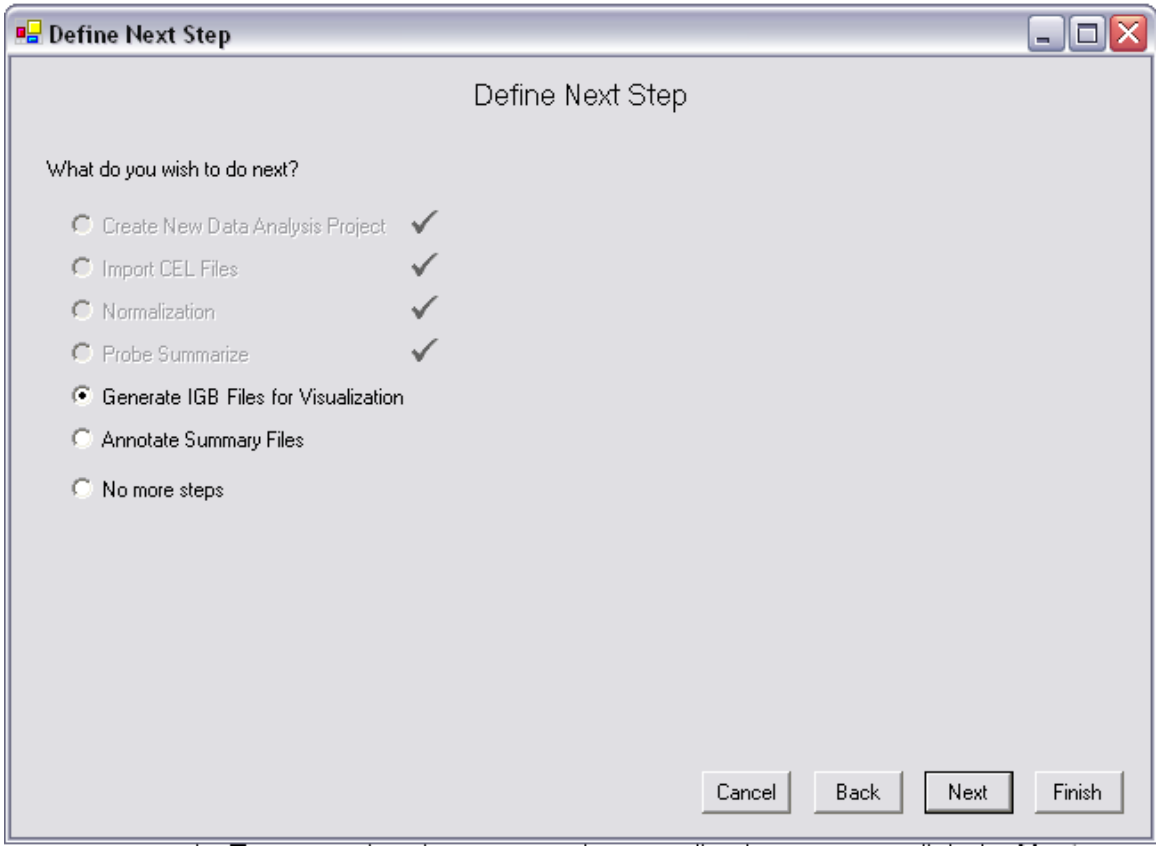

Figure III-8 Define Next Step After Probe Summarize

7 Generate IGB Files for Visualization

The **Generate IGB Files for Visualization** step of an ExACT analysis process allows the user to generate expression graph files (egr) for visualization of exon results within the Integrated Genome Browser (IGB). See Chapter VI: for more information regarding use of egr files within IGB.

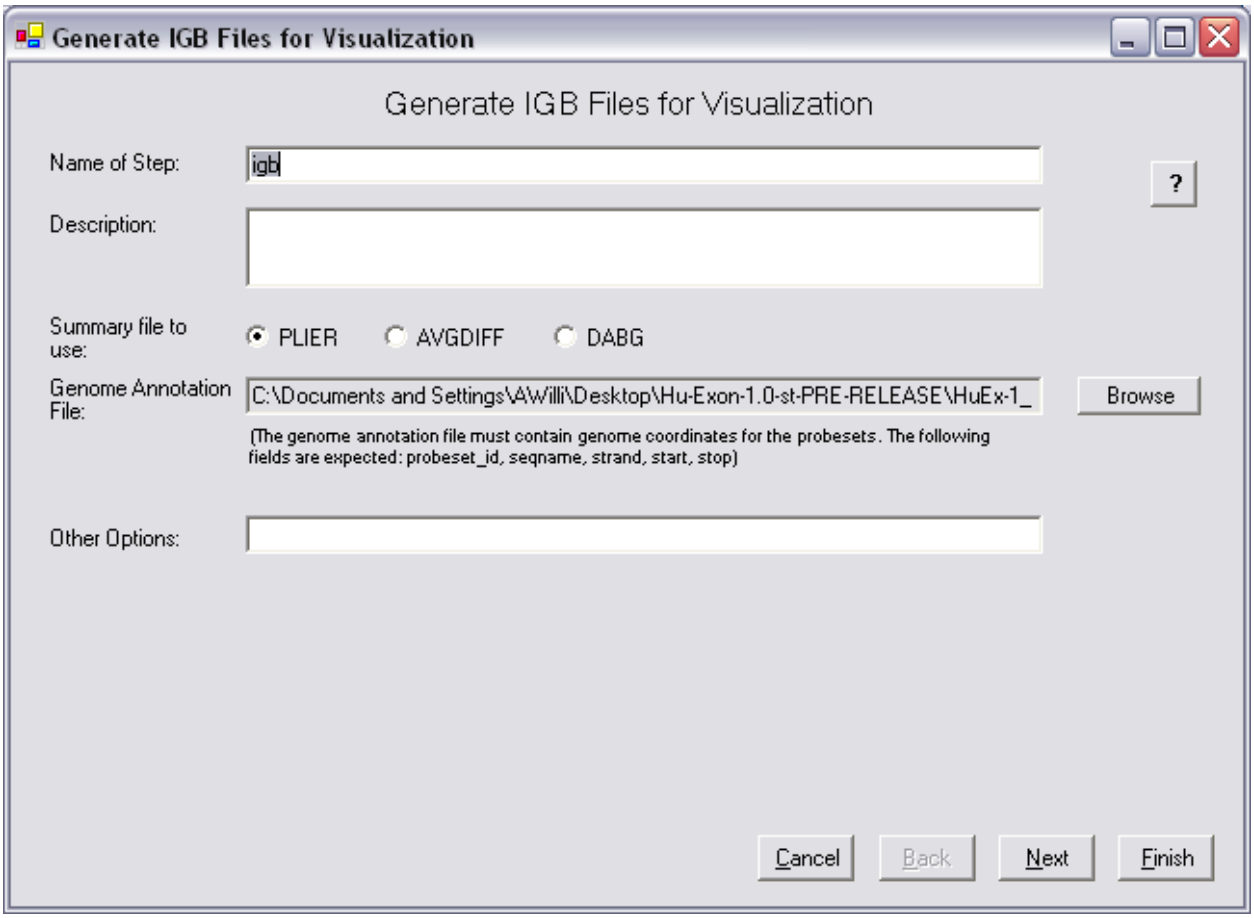

Figure III-9 Wizard Step 5a, Generate IGB Files for Visualization

7.1 Naming the Step

A. Type the name of the step into the **Name of Step** field. Note that the word **igb** is displayed in the **Name of Step** field as a default. The name typed is displayed as shown in Figure III-9, above. This field can be edited by the user for a more detailed step name as required.

B. Note that underscores are automatically entered in the **Name of Step** field when the space bar is keyed.

7.2 Describing the Step (Optional)

A. Type the description of the step being performed into the **Description** field. The description typed is displayed in the **Description** field.

B. Note that spaces, and no underscores, are automatically entered in the **Description** field when the space bar is keyed. The only character count limitation is the size of the field, so very large and detailed descriptions can be entered.

7.3 Selecting Summary File

C. The PLIER, DABG and AvgDiff summary files are available in the **Summary File to Use** area of the **Generate IGB Files for Visualization** dialog box as shown in Figure III-9, above. The user is limited to only selecting a summary file that corresponds to methods selected during the previous probe summarize step.

> **NOTE**: Currently only one summary file type can be selected in this step. Additional visualization steps can be created after the project is executed using other summary files.

D. Click the radio button next to the summary method to be used in creating the egr file.

7.4 Selecting Genome Annotation File

> **NOTE**: The genome annotation file must contain genome coordinates for the probesets. The following fields are expected: probeset_id, seqname, strand, start, and stop. *This is highlighted in the text hint labeled on the Generate IGB File for Visualization dialog box immediately below the field to add a genome annotation file.*
>
> **NOTE**: The HuEx-1_0-st-v2.annot.hg16.csv and the NetAffx exon array probeset CSV files are examples of valid genome annotation files. The particular genome annotation file used should match the array design being analyzed and it should correspond to the genome assembly version (ie UCSC hg16 vs UCSC hg17) that you want to visualize on.

A. Genome annotation files can be selected by clicking the **Add** button to the right of the field.

B. A windows browser directory hierarchy is displayed when the **Add** button is clicked. Select the appropriate genome annotation files from the directories. The path to the folder or file(s) selected is displayed in the field.

7.5 Other Options

Enter any other options as notes to the ExACT user in the **Other Options** field at the bottom of the **Generate IGB Files for Visualization** dialog box. Other options that the user might want to consider are listed in the command line documentation available via the question mark button. One notable advanced option is the **--transcript-cluster-ids** and **--probeset-ids** options which allow the user to generated scored intervals for only those probesets listed in these files.

> **NOTE**: The **Other Options** field is essentially for advanced options. Most users can ignore this field.

7.6 Canceling This Step

A. Click the **Cancel** button on the dialog box to cancel this step.

B. The user is returned to the ExACT main window without saving any analysis step modifications to the project.

7.7 Returning to the Previous Step

A. Click the **Back** button on the dialog box to return to the previous step.

B. The previous step in the ExACT process is displayed

7.8 Define Next Step After Generate IGB File for Visualization

The **Define Next Step** dialog box is displayed by clicking the **Next** button on the **Generate IGB File for Visualization** dialog box. This dialog box at this stage as shown in Figure III-10, below. Note that the **Annotate Summary Files** radio button, the next step, has been selected.

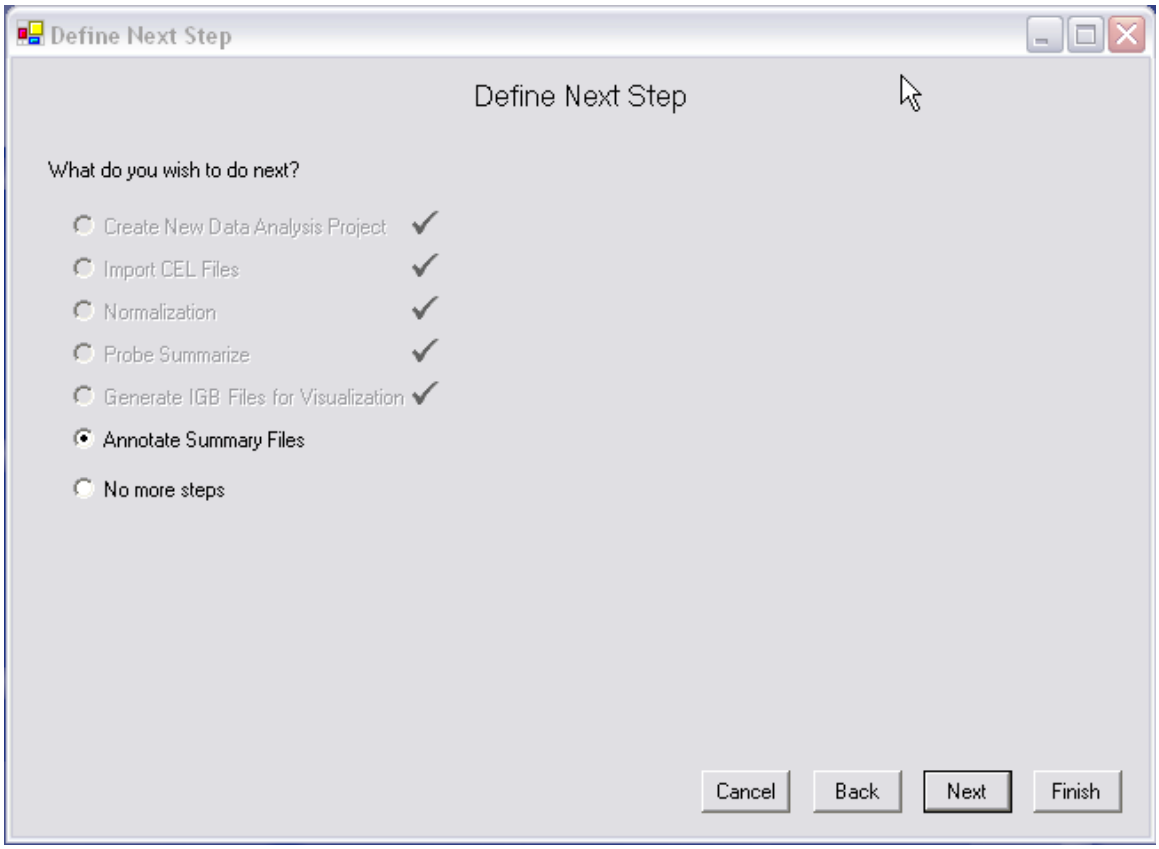

Figure III-10.. Define Next Step After Generate IGB Files for Visualization

8 Annotate Summary Files

The **Annotate Summary Files** segment of this ExACT process allows the user to merge textual annotations (such as the NetAffx probeset CSV annotation file) with the summary files.

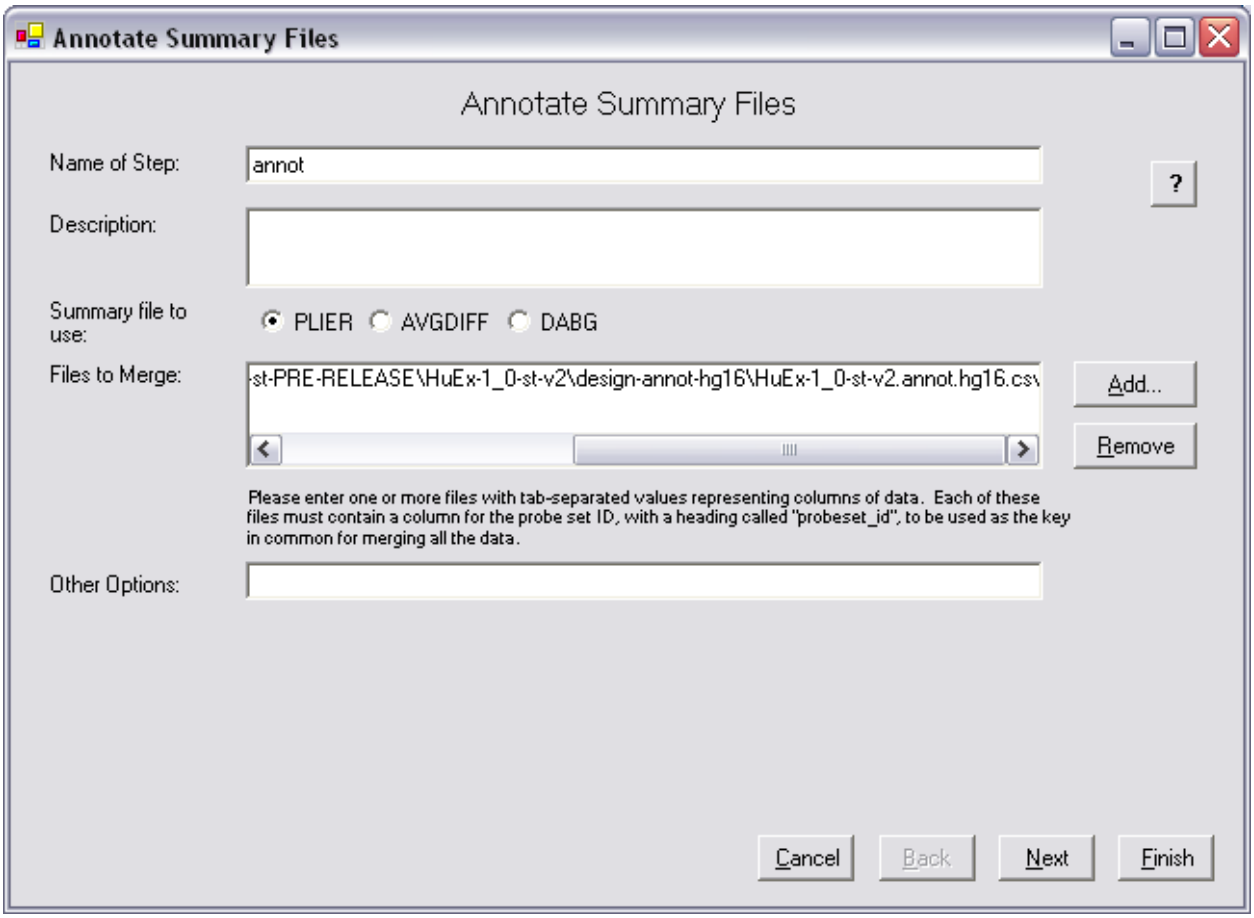

Figure III-11 Wizard Step 5b, Annotate Summary Files

8.1 Naming the Step

A. Type the name of the step into the **Name of Step** field. Note that the word **annot** is displayed in the **Name of Step** field as a default. The name typed is displayed as shown in Figure III-11, above. This field can be edited by the user for a more detailed step name as required.

B. Note that underscores are automatically entered in the **Name of Step** field when the space bar is keyed.

8.2 Describing the Step (Optional)

A. Type the description of the step being performed into the **Description** field. The description typed is displayed in the **Description** field.

B. Note that spaces, and no underscores, are automatically entered in the **Description** field when the space bar is keyed. The only character count limitation is the size of the field, so very large and detailed descriptions can be entered.

8.3 Selecting Summary Files

A. The PLIER, DABG and AvgDiff summary files are available in the **Summary File to Use** area of the **Annotate Summary Files** dialog box as shown in Figure III-11, above. The user is limited to only selecting a summary file that corresponds to methods selected during the previous probe summarize step.

B. Click the radio button of the summary file to annotate.

> **NOTE**: Currently only one summary file type can be selected in this step. Additional visualization steps can be created after the project is executed for the other summary files.

8.4 Selecting Files to Merge

A. To select the summary files for this project. Click the **Add** button to the right of the **Files to Merge** field.

> **NOTE**: Annotation files (tab separated TSV or comma separated CSV files) may be merged with ExACT analysis results. For example, the same genome annotation files used to create egr files during the previous step may be used for this step as well. Unlike genome position files, these annotation files are only required to have a probeset_id column.

B. A windows browser directory hierarchy is displayed when the **Add** button is clicked. Select the appropriate files to be merged from the directories. The path to the file(s) selected is displayed in the field.

C. To remove a file(s) from the **Files to Merge** field, select the file to be deleted in the field. Click the **Remove** button to the right of the field. The selected file is removed.

8.5 Other Options

Enter any other options as notes to the ExACT user in the **Other Options** field at the bottom of the **Annotate Summary** dialog box.

> **NOTE**: The **Other Options** field is essentially for advanced options. Most users can ignore this field.

8.6 Canceling This Step

A. Click the **Cancel** button on the dialog box to cancel this step.

B. The user is returned to the ExACT main window without saving any analysis step modifications to the project.

8.7 Returning to the Previous Step

A. Click the **Back** button on the dialog box to return to the previous step.

B. The previous step in the ExACT process is displayed.

8.8 Define Next Step After Annotate Summary Files

The **Define Next Step** dialog box is displayed after clicking the **Next** button on the **Annotate Summary Files** dialog box. This dialog box at this stage as shown in Figure III-12, below.

> **NOTE**: This is the finishing point of this ExACT analysis project. At this point, all the check marks are displayed next to all steps in the ExACT process, the Next button is shown disabled, and the radio button next to No More Steps is selected.

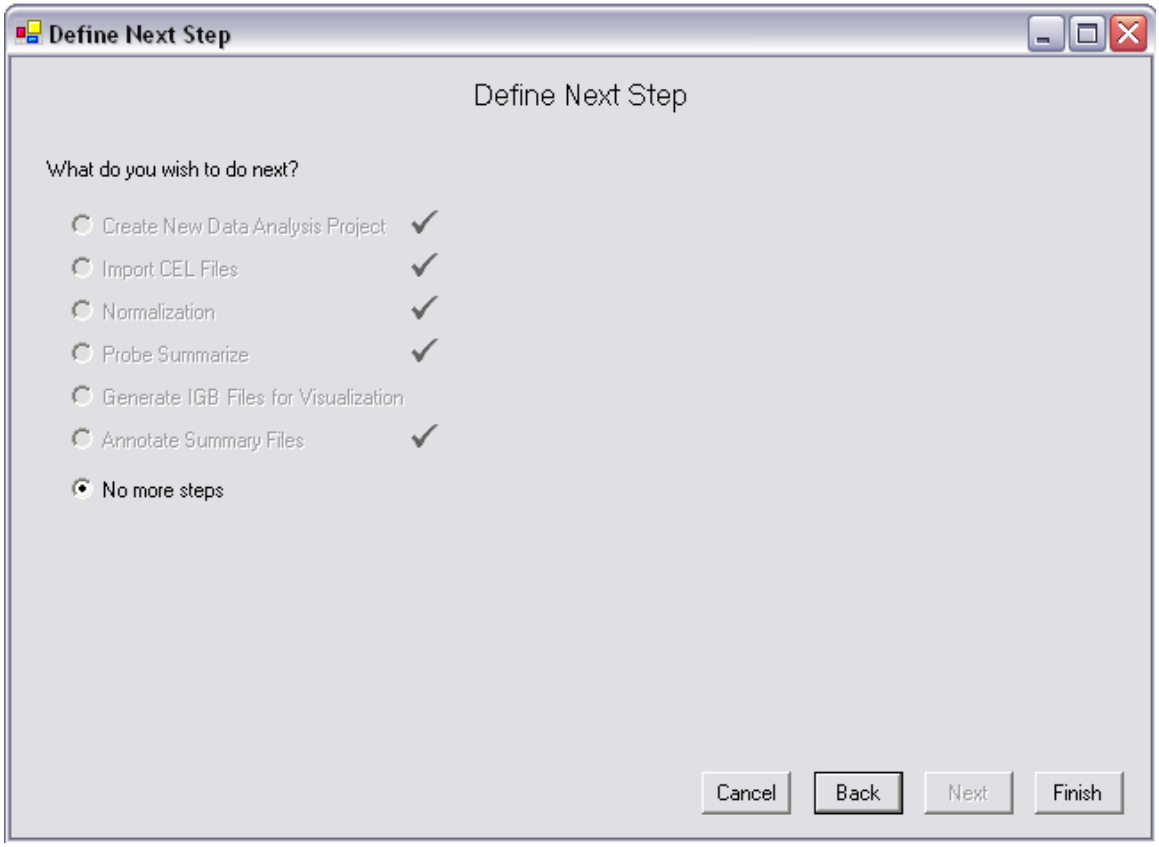

Figure III-12 Define Next Step After Annotate Summary Files

Chapter IV: Executing an ExACT Analysis

This chapter covers the **Progress of Execution** dialog box.

1 Progress of Execution Window

The **Progress of Execution** dialog box is displayed as shown in Figure IV-1, below, when the **Finish** button is clicked. This dialog box will be replaced by the main window once all of the analysis steps are complete. The particular analysis step being executed is shown near the top as seen in Figure IV-1 below for the normalization step. Information and error messages for each of the analysis step command line programs will be displayed in the main next window of the **Progress of Execution** dialog box as shown in Figure IV-1 for the normalization step

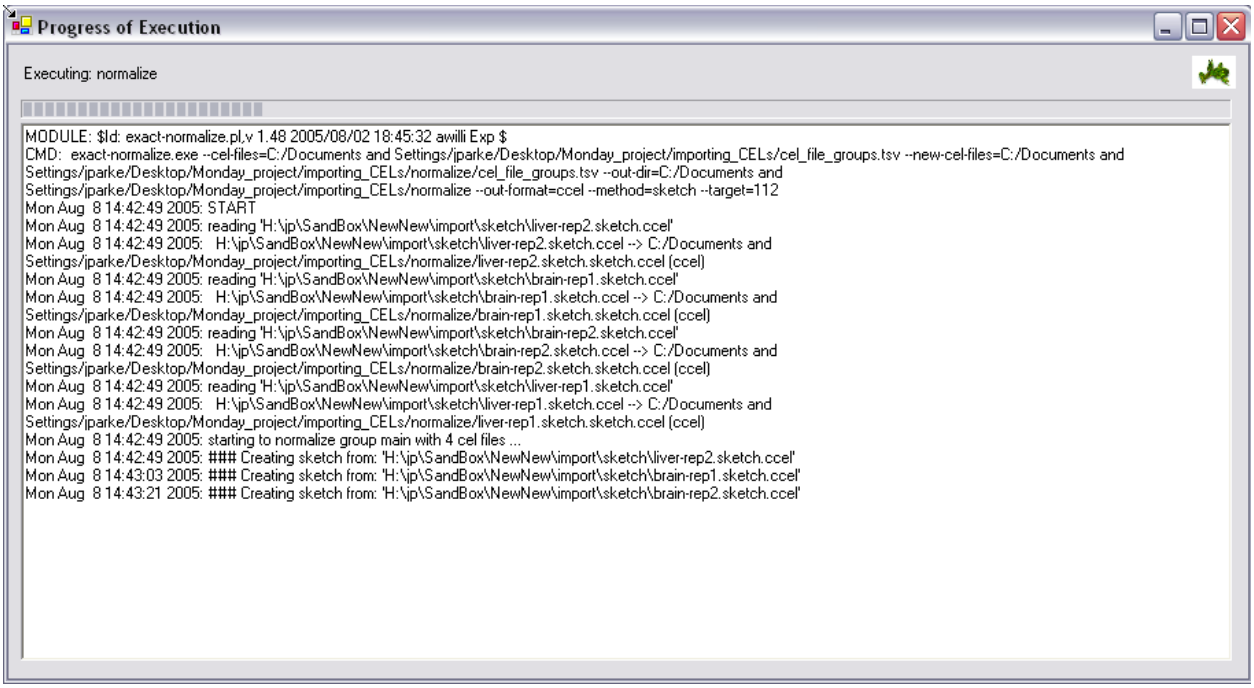

Figure IV-1 Progress of Execution Dialog Box

> **NOTE**: In rare instances, the log information displayed will not be updated when the next step is executed. For example, the log display may indicate that the normalization step successfully completed and no log information is displayed for the probe summarization step. The user should allow the analysis to complete and the probe summarization log will be displayed in the Results Log tab after execution is complete. You can confirm which ExACT command line program is currently running by looking for "exact-*" processes under the windows task manager (accessible using CTRL-ALT-DEL).

2 Cancel Execution

Follow the procedures listed below to cancel Progress of Execution. Specifically, there is no facility within the ExACT GUI Workflow to cancel the command line analysis program being executed. So if the user wants to cancel a running analysis they must use the windows task manager as outlined here.

A. To cancel an ongoing progress of execution, simultaneously hold down the **Ctrl**+ **Alt** + **Delete** keys. The **Windows Security** dialog box is displayed.

B. Click the **Task Manager** button. The **Windows Task Manager** dialog box is displayed.

C. Click the **Processes** tab. The list of processes is displayed in the active window.

D. Locate the ExACT process on the **Processes** tab that is currently active by looking for the "exact-*" process in the **Image Name** list. When sorted by % CPU utilization, it should be listed at or near the top of the list.

E. Highlight the ExACT process and click the **End Process** button at the bottom of the dialog box. Deletion of the ExACT process from the **Processes** tab confirms that the analysis has been stopped.

3 Log Messages

The Results Log messages are displayed in the **Results** tab of the main window. The messages in the Results Log display various kinds of information relevant to the ExACT analysis step and are generated by the command line program for that step. The information includes timestamps for steps performed and command line parameters executed per the choices made in the corresponding **Workflow Wizard** step configuration dialog box.

> **NOTE**: The Results Log for each step is saved in the same folder as the results for that step in a file called report.txt. Please include this file when communicating execution problems to Affymetrix support personnel.

Error messages and warnings will also be displayed. Critical **Results Log** error messages are displayed in capital letters. For example, a **FATAL** error will be labeled as such in the log, and indicate the reason **Progress of Execution** stopped at that point.

> **NOTE**: Always review the results log for each of the steps to ensure that there were no problems, and that each step reported a successful completion by displaying the "**`EXIT: 0 (success)`**" label at the end of the log report.

Chapter V: ExACT Workflow Wizard Main Window with Open Project

1 Introduction

Once a project has been created and executed, or a previously run project has been opened, the user can use the Main Window for a variety of tasks including:

- View the settings used to run the analysis
- Review the report log for each analysis step
- Open results folders
- Create new steps
- Clone existing steps
- Delete steps

After the **Progress of Execution** process has been completed (or a previously run project has been opened), the ExACT main window will be displayed as shown in Figure V-1, below. See Chapter II: , for more information about the main window when there is no active project.

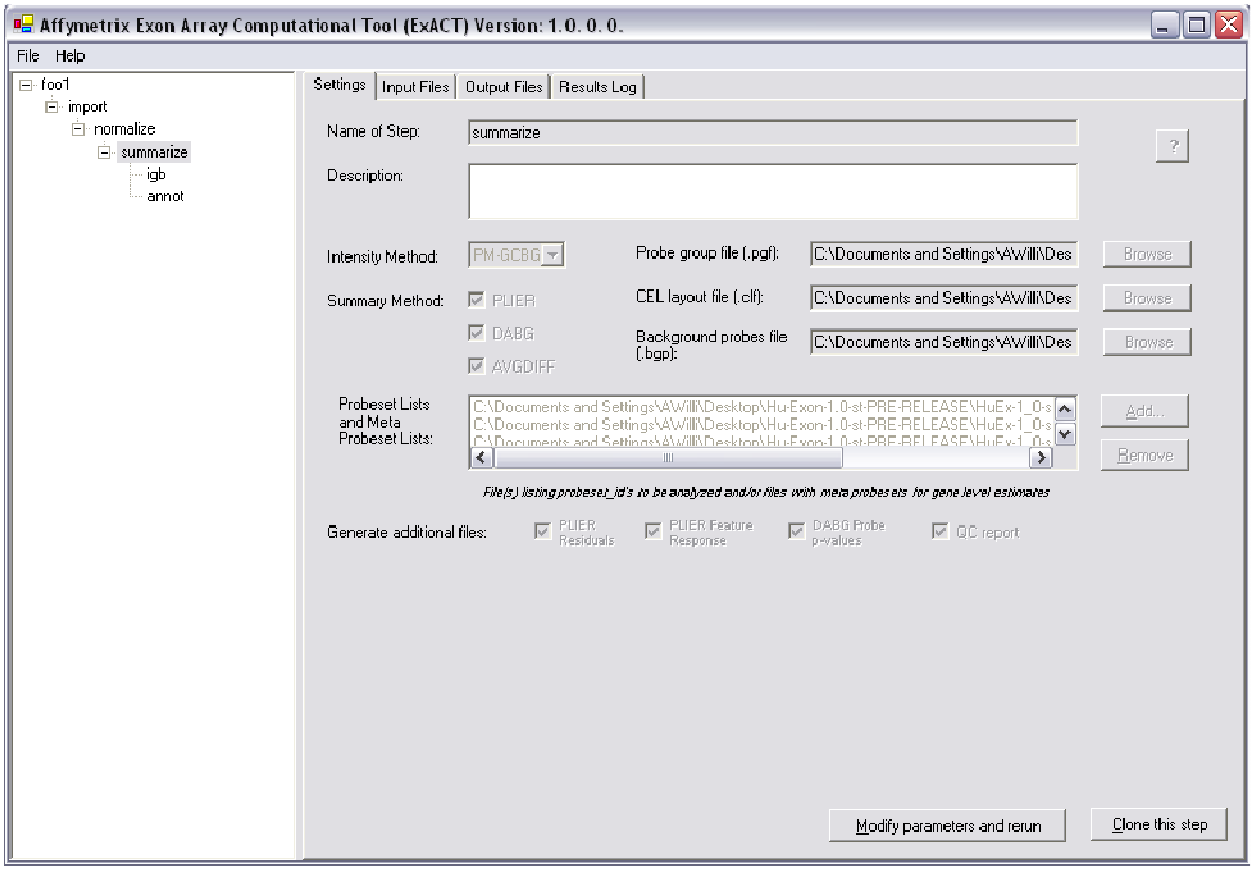

Figure V-1: ExACT Main Window with Open Project

2 Project Tree View

The window on the left in Figure V-1, above, displays the project tree view for this ExACT project. The steps completed to this point will be nested in the tree relative to their interdependencies. From the project tree, new steps can be created, and existing steps can be removed by clicking the right mouse button on a particular step as shown in Figure V-2, below.

To view the entire directory hierarchy in the left pane, if it is hidden by the left active window, direct the pointer to the vertical split bar dividing the two panes of the dialog box. Hold the left mouse button down and drag the split bar to the left until the entire directory tree in the left window is displayed.

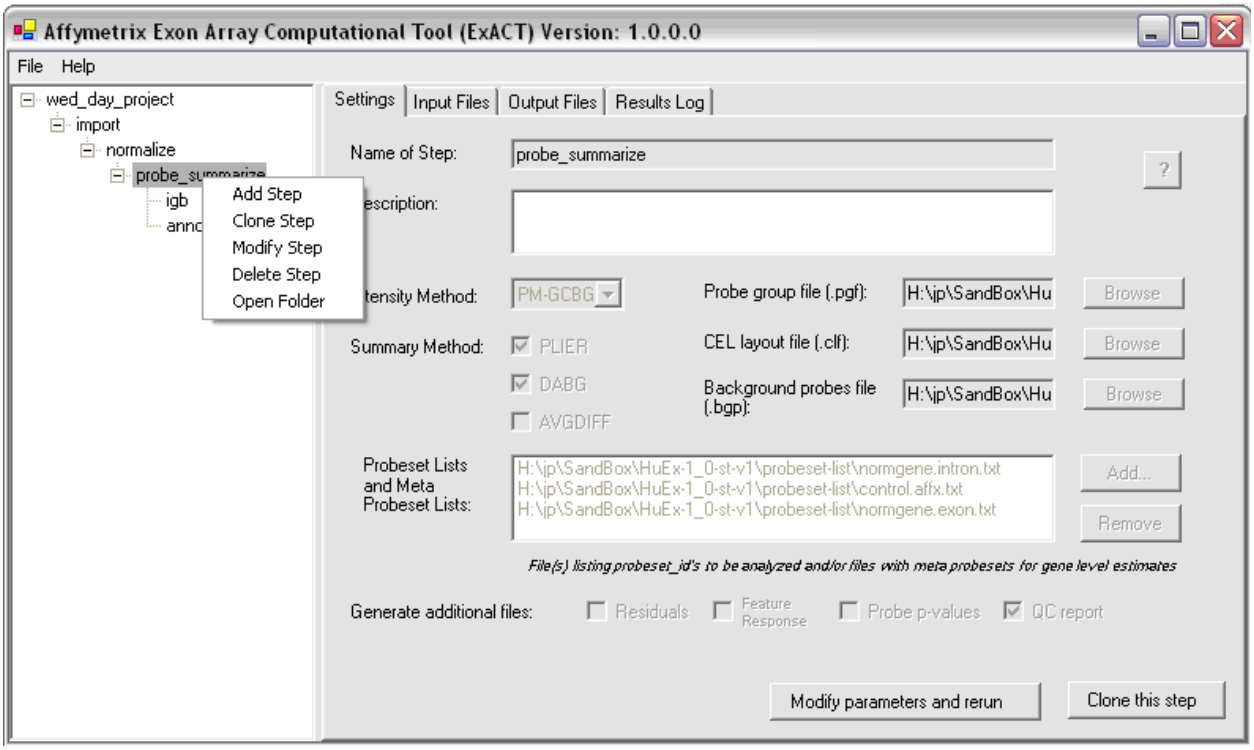

Figure V-2: Additional Steps from the Project Tree View on Main Window

2.1 Add Step

Select this particular option to add a new step. The new step will follow the currently selected step. Selection of this menu option will cause the Define Next Step dialog box from the Workflow Wizard (Figure III-4) to be displayed. See Chapter III: for more information on using the Workflow Wizard.

2.2 Clone Step

Select this particular option to create a copy of an existing step. This menu option is useful when the user wants to rerun a particular analysis step with different parameters. When selected, the user will be dropped into the particular Workflow Wizard dialog box for that particular analysis step. See Chapter III: for more information on using the Workflow Wizard.

2.3 Modify Step

Select this particular option to modify an existing step. Note that when the user selects this option the existing results and all subsequent analysis steps in the tree will be removed. Because of this users should consider using the Clone Step option over this one. If the user selects the Modify Step option, the following dialog box, Figure V-3, will be displayed warning the user. The appropriate Workflow Wizard Window will be displayed if the user chooses to modify the parameters and re-run.

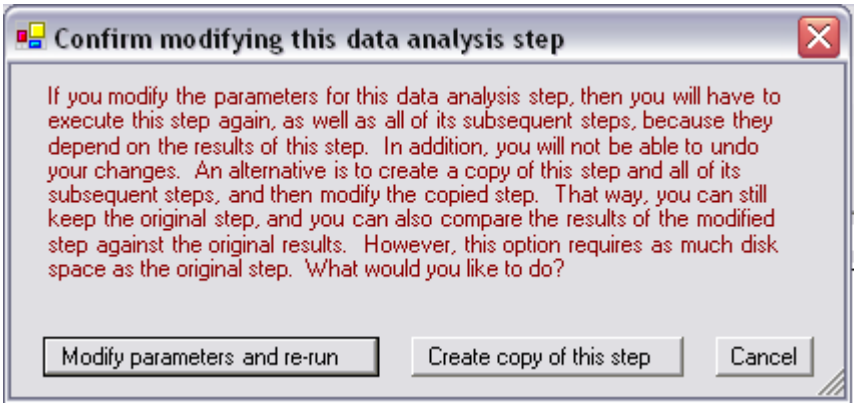

Figure V-3: Confirm modifying this data analysis step Dialog Box

2.4 Delete Step

Selecting this menu option will cause the current analysis step and all subsequent analysis steps to be removed. All data generated by those steps will also be removed.

2.5 Open Folder

This option will open an MS Windows Explorer window of the results folder. Using this option the user can gain access to the files generated by this step.

> **CAUTION**: Modification of the files is not recommended. Altering or removing files may result in unexpected failures when altering or adding to the project using the Workflow Wizard.

3 Project Information Tabs

The project information tabs contain information about the currently selected analysis step. Information about the settings used to run that analysis step, files used as input by the analysis step, files generated by that analysis step, and the Report Log from that analysis step are all accessible from the project information tabs.

3.1 Settings Tab

The settings used for a particular analysis step can be viewed by clicking on the **Settings** tab. The specific information displayed will depend on the particular analysis step which is selected. An example is shown in Figure V-1, above.

NOTE: The Settings tab contains the same data as the corresponding Workflow Wizard step configuration dialog box. All the settings are disabled which can make it hard to view the setting information when some of the field contents are larger than the field window (ie really long file names that do not fit within the visible text box). To view the entire contents, select Clone Step which will bring up an active configuration window. When the user is done checking the settings, select cancel, which will abort the creation of a new step.

3.2 Input Files Tab

Clicking on the **Input Files** tab will display the files used as input for that particular analysis step. An example of this is shown below (Figure V-4, below) where the normalized CEL files used as input to the PLIER summarization step are listed.

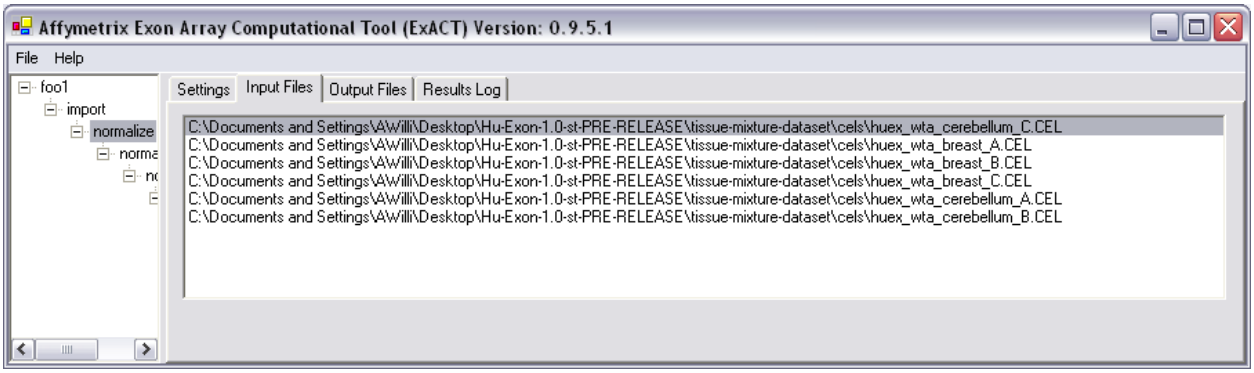

Figure V-4 : Input Files Tab

3.3 Output Files Tab

Click on the **Output Files** tab will display the files generated by that particular analysis step. An example is shown below (Figure V-5) where the output files generated by the PLIER summarization step are displayed. The user can open and copy these files by selecting the Open Folder option. (See Chapter V: 2.5 for more information.)

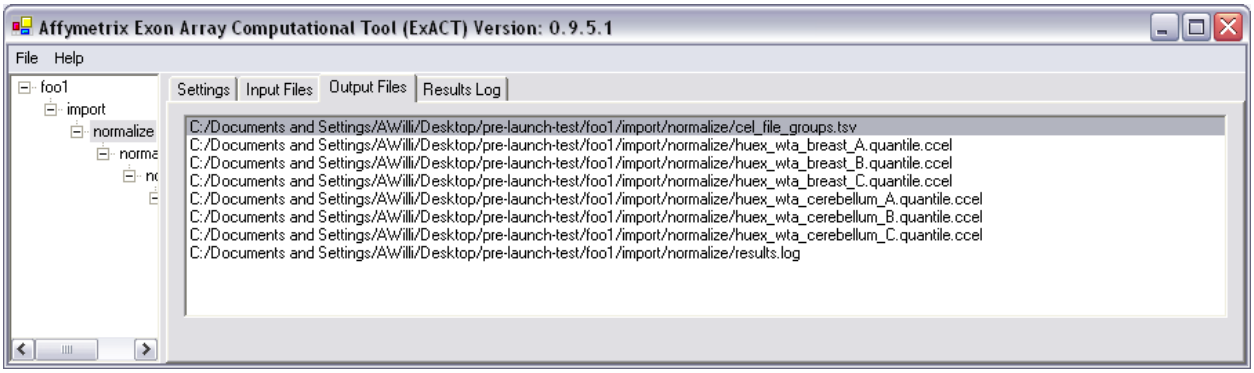

Figure V-5 Output Files Tab

3.4 Results Log Tab

Click on the **Results Log** tab to view information generated during the execution of that analysis step. A successfully completed step should end with the line "**`EXIT: 0 (Success)`**". An example of a

successfully completed analysis step is shown below (Figure V-6). Use the **Save Results Log** button to export the results log.

> **NOTE**: When contacting support@affymetrix.com regarding problems with a particular ExACT step, the results log should be included for reference.

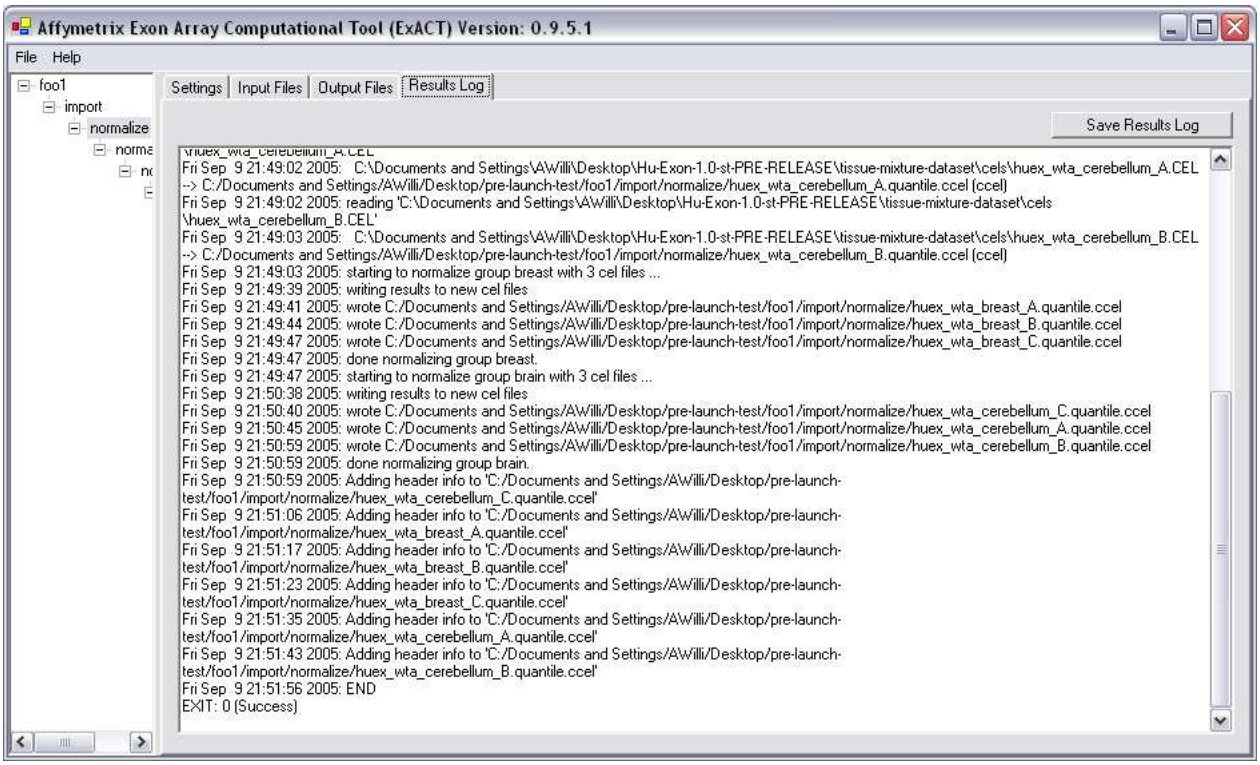

Figure V-6: Results Log Tab

Chapter VI: Importing ExACT Results into IGB

1 Introduction

The Integrated Genome Browser (IGB) can be used to visualize numerical results from ExACT and other software in a genome context as shown in Figure VI-1. This is accomplished by loading the egr files generated by the ExACT IGB Visualization step into IGB.

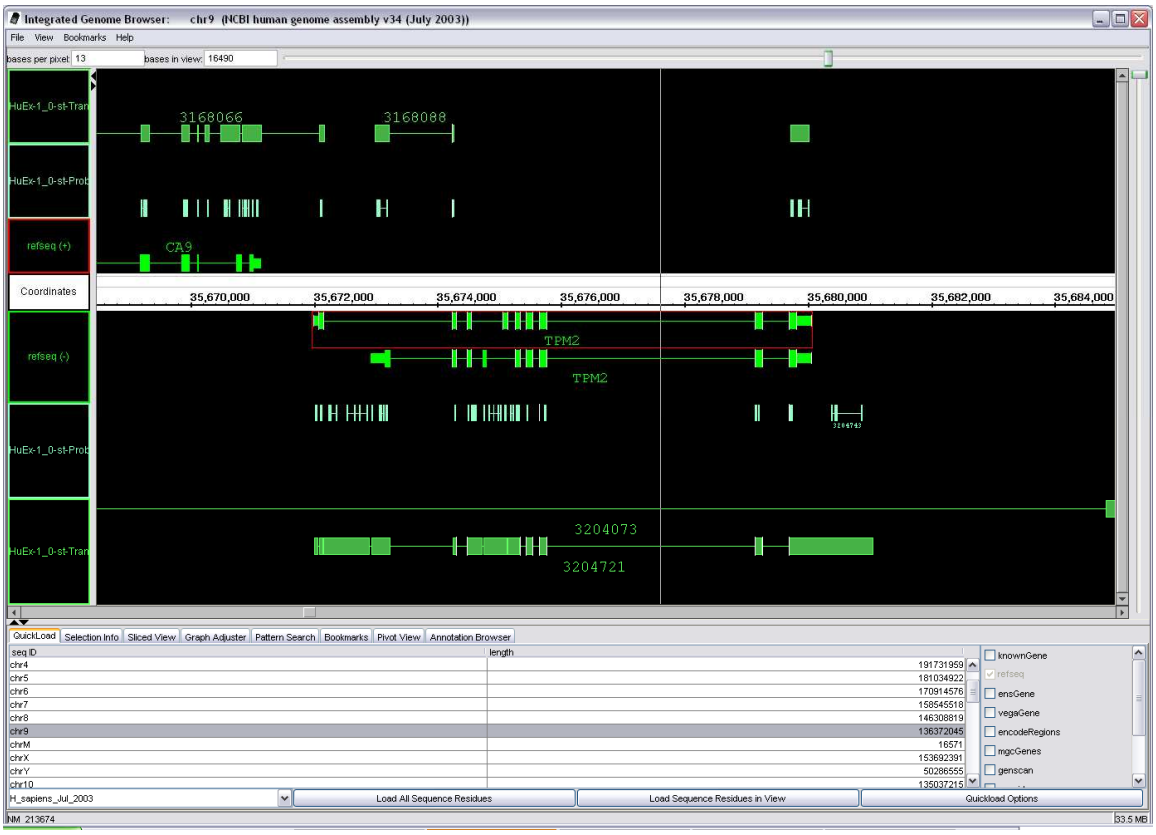

Figure VI-1: Integrated Genome Browser (IGB) Main Window

There is a separate IGB manual which covers the general aspects of using IGB, and this section assumes that the user is familiar with the contents of the IGB manual. See the **Integrated Genome Browser User's Guide** and the IGB product page at http://www.affymetrix.com/support/developer/tools/download_igb.affx for specific information.

2 Starting IGB

IGB is a Java application. The simplest way to run IGB is to use the Java Web Start feature which is installed with newer versions of Java. The user must have Java installed (see J2SE download from http://java.sun.com/j2se/index.jsp).

When starting IGB from the Affymetrix website using Web Start, the user must select a specific memory configuration. For example, there may be an IGB_512 and IGB_756 configuration, where the first will use at most 512 Mb, and the second will use at most 756 Mb. The user should choose the configuration with the largest memory limit which is still less than the total capacity of the physical memory installed on the user's computer.

NOTE: After running IGB once using Web Start, there will be an IGB program listed by the Web Start application. Using the Web Start application under Start->All Programs->Java Web Start, the user can enable the console log for IGB. To do this, select the IGB application and then File->Preferences. On the Advanced tab check the Show Java Console option. Messages on the console may be helpful in trouble shooting IGB problems including out of memory errors. (See Figure VI-2.)

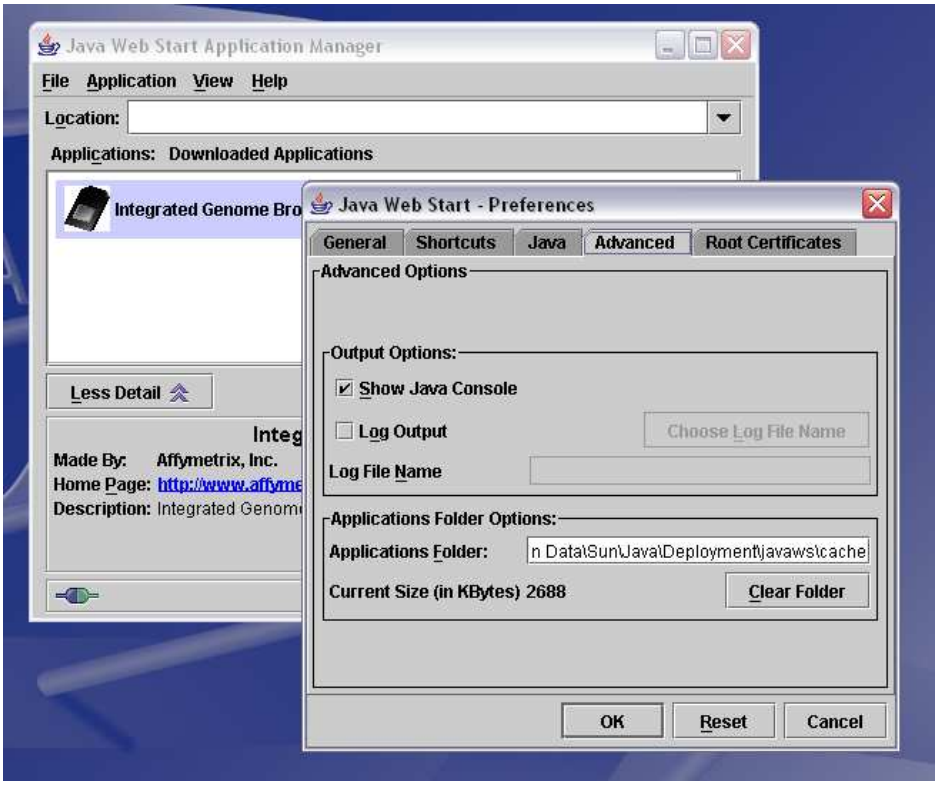

Figure VI-2: Java Web Start and Enabling Java Console Output

3 Loading a Particular Genome

Once IGB has started the user should go to the **QuickLoad** tab and load the appropriate genome as shown in Figure VI-3. The genome version used in the genome position file when the Expression Graph (egr) file was created should match the genome version loaded here. After the genome is loaded, the user must select the chromosome of interest on the quick load tab. Data will be displayed in the main genome viewer window and will probably display RefSeq annotations by default. Additional annotations may be loaded using the **QuickLoad** tab by checking additional annotation sources on the right hand side of the tab. Additional annotations may also be loaded from a DAS server using the **File**->**Load DAS Features**.

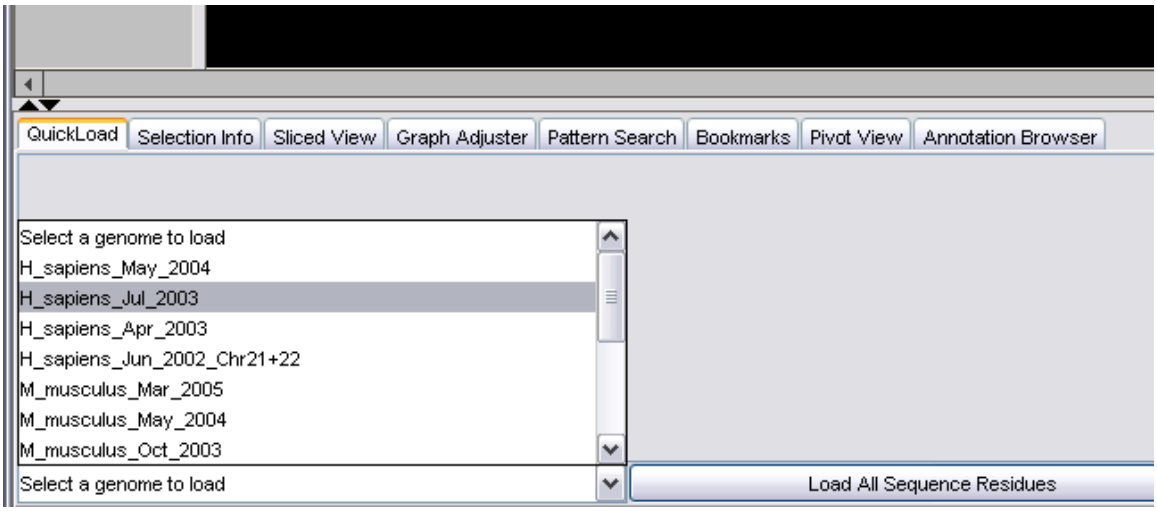

Figure VI-3: Loading a Genome Into IGB

4 Loading the Chip Design

The next step is to load the probeset information for the particular CEL files which were analyzed in this ExACT analysis.

This step is optional. There are two options for loading these annotations. If the user is only interested in specific gene regions, then the user should use the DAS approach. If the user wants to view whole or large portions of chromosomes then the BGN approach should be used.

4.1 Loading from a Distributed Annotation Server (DAS)

Transcript cluster and probeset annotations can be loaded from the Affymetrix NetAffx DAS server. This DAS server should be displayed in the default list of DAS servers. The user will first need to select a particular gene region of interest. (See the ***IGB User's Guide*** for more information on navigating and searching in IGB.)

Once that region has been located, the transcript cluster and probeset annotations can be loaded using the **File**->**Load DAS Features** menu option.

Select the NetAffx DAS server and then check the transcript and probeset annotation tracts. Then click **OK** to load those annotations. Once loaded, the user should be able to link back to NetAffx by right clicking on a transcript or probeset annotation and selecting the additional information tab.

> **NOTE**: Users can use the NetAffx website as a shortcut to this. Search on NetAffx for a particular probeset or transcript cluster of interest and then use the NetAffx links on the transcript cluster or probeset information page to start up IGB. The genome annotations including transcript and probeset annotations will be automatically loaded. The user can then zoom out and do the DAS load described above to load additional flanking annotations. One caveat with this approach is that the genome position file used to create the egr files must correspond to the current NetAffx genome version. The NetAffx genome version is indicated near the IGB links within the NetAffx Analysis Center.

4.2 Loading from Binary Annotation Files (BGN)

To visualize annotations for entire chromosomes, the user should download the zip files from the Affymetrix website containing the BGN files for the genome version of interest. The user can then load

transcript and/or probeset annotations on a per chromosome basis. Links back to NetAffx should still function as described for DAS-loaded annotations.

4.3 Loading the Expression Graph (egr) File

To load the egr file generated by ExACT, use **File**->**Open** menu option and then browse to the appropriate egr file. When loaded, graph tracts will appear showing the probeset results for each array analyzed as seen in Figure VI-4 for a localized region around the TPM2 gene. Use the graph adjuster tab to adjust how the ExACT egr results are displayed.

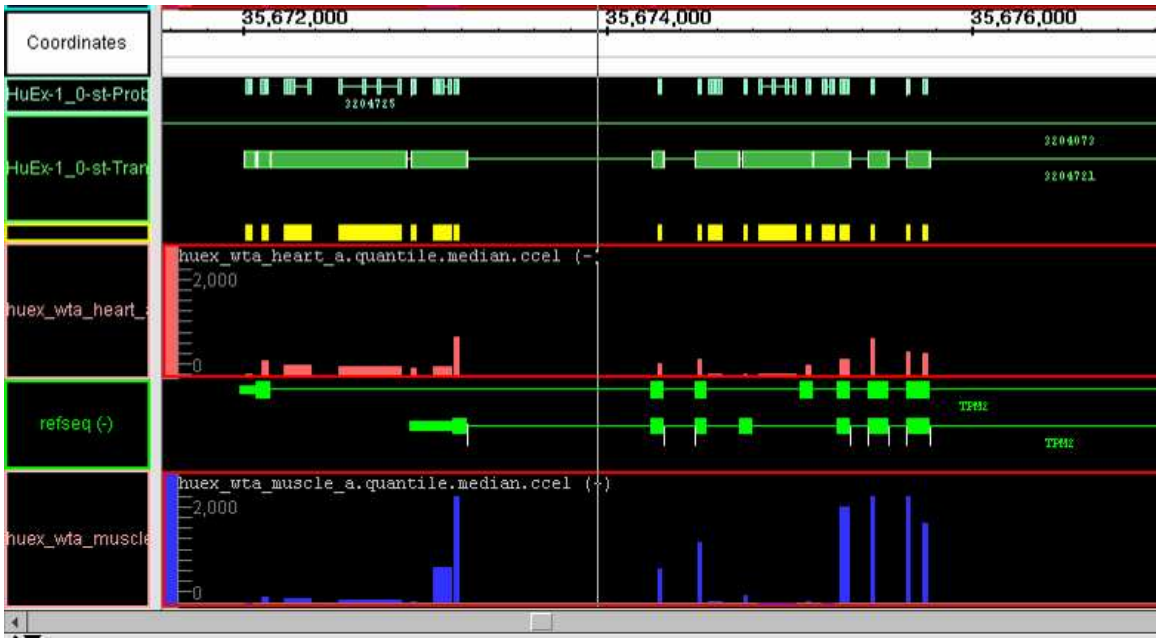

Figure VI-4: IGB Graphs Loaded from egr Files

The tract with yellow annotations in Figure VI-4 reflects the probeset regions reported in the egr file. By selecting multiple annotations on the yellow (in this example) tract, ExACT results will be reported in the pivot view shown in Figure VI-5.

> **NOTE**: By selecting various annotations such as RefSeq or Ensembl transcripts along with the probeset annotations in the egr tier, a full transcript annotation will show up in the pivot view (as seen in Figure VI-5).

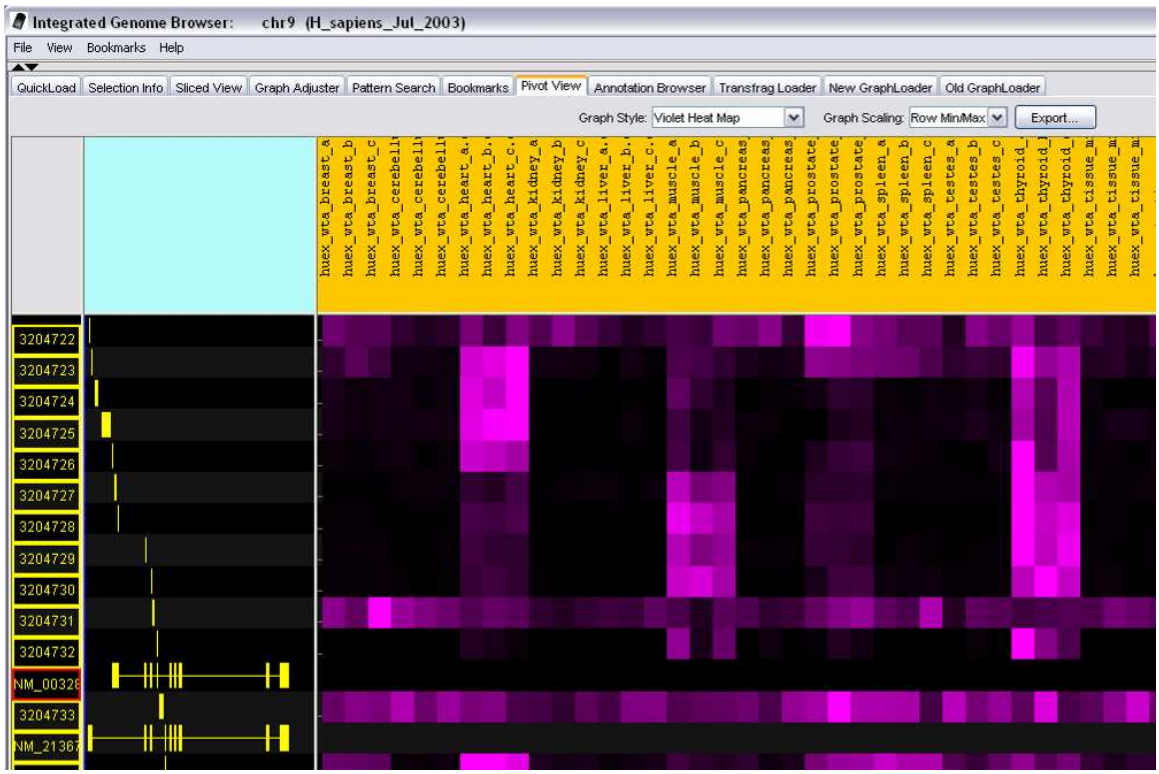

Figure VI-5: IGB Pivot View Loaded from egr Files

Appendix I: ExACT Software License

Affymetrix, Inc. Client Software (No Charge)
End User License Agreement

IMPORTANT-READ CAREFULLY: This End-User License Agreement ("***Agreement***") is a legal agreement between you and Affymetrix, Inc. ("***Affymetrix***"), the licensor of the software ("***Software***") and related media, printed materials and/or electronic documentation ("***Documentation***") (together, "***Software Products***") provided to you in connection with this Agreement. The Software Products are for use on a computer system (***"System"***) that strictly complies with specifications set forth in the Documentation. BY CLICKING ON THE "ACCEPT" BUTTON DURING THE INSTALLATION PROCESS, OPENING THE PACKAGE CONTAINING THE SOFTWARE PRODUCTS, OR OTHERWISE USING THE SOFTWARE PRODUCTS, YOU AGREE TO BE BOUND BY THE TERMS OF THIS AGREEMENT. IF YOU DO NOT AGREE TO THE TERMS OF THIS AGREEMENT, AFFYMETRIX IS UNWILLING TO LICENSE THE SOFTWARE PRODUCTS TO YOU. IN SUCH EVENT, YOU SHALL NOT USE, READ OR ACCESS IN ANY MANNER THE SOFTWARE PRODUCTS, AND YOU SHOULD (A) PROMPTLY CONTACT AFFYMETRIX OR THE DISTRIBUTOR WHO PROVIDED THE SOFTWARE PRODUCTS TO YOU ("***DISTRIBUTOR***") WITHIN THIRTY (30) DAYS OF FIRST ACQUIRING THE SOFTWARE PRODUCTS FOR INSTRUCTIONS ON RETURN OF THE SOFTWARE PRODUCTS, OR (B) CLICK THE "CANCEL" BUTTON AND THE INSTALLATION PROCESS WILL NOT CONTINUE. WRITTEN APPROVAL IS NOT A PREREQUISITE TO THE VALIDITY OR ENFORCEABILITY OF THIS AGREEMENT AND NO SOLICITATION OF ANY SUCH WRITTEN APPROVAL BY OR ON BEHALF OF AFFYMETRIX SHALL BE CONSTRUED AS AN INFERENCE TO THE CONTRARY. IF YOU HAVE ORDERED THE SOFTWARE PRODUCTS, AFFYMETRIX'S ACCEPTANCE IS EXPRESSLY CONDITIONAL ON YOUR AGREEMENT TO THESE TERMS TO THE EXCLUSION OF ALL OTHER TERMS; IF THESE TERMS ARE CONSIDERED AN OFFER BY AFFYMETRIX, ACCEPTANCE IS EXPRESSLY LIMITED TO THESE TERMS.

1. Software License. Subject to the terms and conditions of this Agreement, Affymetrix grants to you only, and you accept, a nonexclusive, nontransferable, nonsublicensable license ("***License***") to install and use the Software Products in conjunction with other Affymetrix products for purposes described in the Documentation, *provided that* the Software shall only be used (a) on the System (b) in accordance with the Documentation, and (c) in machine-readable, object code form. You shall use the Software solely for your internal operations and Affymetrix retains all rights not expressly licensed herein.

2. Restrictions on Use. (a) You will not modify, incorporate into or with other software, or create derivative works of the Software Products. (b) You may not use the Software Products, except as set forth in Section 1. Any copies that you are permitted to make pursuant to this Agreement must contain the same copyright and other proprietary notices that appear on or in the Software Products. (c) You will not reverse engineer, decompile, disassemble, or otherwise translate or analyze the Software, any source code, underlying ideas, algorithms or programming by any means whatsoever. (d) The Software Products are licensed to you for research use only. You agree not to use the Software Products in any setting requiring FDA or similar regulatory approval. (e) You will not transmit or use the Software Products over a network (except for Software that accesses or otherwise uses the services of or databases created by or related to Affymetrix server-based software, and then only in accordance with the Documentation and applicable Affymetrix end user license agreement and documentation for such Affymetrix server-based software) or disseminate performance information or analysis to any third party relating to the Software Products. (f) You agree not to remove any product identification, copyright or other notices. (g) You agree not to load or use any portion of the Software Products on or with any machine or system other than the System.

3. Intellectual Property. (a) The Software Products and any copies that you are authorized by Affymetrix to make are the intellectual property of and are owned by Affymetrix and protected under copyright and other intellectual property laws (including international copyright treaties, as well as other intellectual property treaties). (b) You acknowledge and agree that all right, title, and interest in and to the Software Products, including associated intellectual property rights, are and shall remain with Affymetrix. (c) This Agreement does not convey to you an interest in or to the Software Products, but only a limited right of use which is revocable in accordance with the terms of this Agreement. (d) The

Software Products are licensed, not sold, to you by Affymetrix for use only under the terms of this Agreement, and Affymetrix reserves any rights not expressly granted to you. (e) You shall affix, to each full or partial copy of Software Products made by you, all copyright and proprietary information notices as appeared on or in the original. (f) No license, right, or interest in any Affymetrix trademark, trade name, service mark, or other designation is granted hereunder. (g) Nothing contained herein shall be construed as conferring by implication, estoppel or otherwise any license or right under any patent, whether or not the exercise of any right herein granted necessarily employs an invention of any existing or later issued patent. Notwithstanding the above, Affymetrix agrees that you shall have the right to exercise all rights expressly granted to you pursuant to this Agreement. (h) Except as expressly stated above, this Agreement does not grant you any intellectual property rights in the Software Products. (i) You acknowledge and agree that any feedback, suggestions, comments, improvements, modifications and other information that you provide to Affymetrix relating to the Software Products or their performance may be used, disclosed, disseminated and/or published by Affymetrix for any purpose, including incorporating such information in improvements to the Software, without obligation of any kind to you, and that you waive any rights whatsoever in or to such information.

4. Transfer and Other Uses. (a) You will not assign, sublicense, transfer, pledge, distribute, market, lease, loan or rent to any third party the Software Products or use the Software Products in any service bureau, time sharing or rental arrangement, share your rights under this Agreement, engage in the service of operating the Software on behalf of third parties, or otherwise use or allow others to use Software Products to or for the benefit of third parties or authorize all or any portion of the Software Products to be copied onto another user's computer except as may be expressly permitted herein. (b) You may, however, load the Software on the System in both temporary and permanent storage, and the associated storage media may be utilized for backup purposes. In addition, you may make one copy of the Software on a backup storage medium for the purpose of backup.

5. Term and Termination. (a) This Agreement shall continue until terminated. (b) You may terminate this Agreement at any time by destroying all copies of the Software Products, including the associated storage media and Documentation, and erasing all copies of the Software Products in both temporary and permanent storage on the System. (c) Affymetrix may terminate this Agreement immediately upon the breach by you of any term hereof. (d) Upon termination, you shall immediately cease all use of the Software Products and return or destroy all copies of the Software Products and all portions thereof and so certify the same to Affymetrix. Termination is not an exclusive remedy and all other remedies will be available to Affymetrix whether or not the Agreement is terminated.

6. No Warranty. The Software Products are provided "AS IS" and Affymetrix makes no warranty as to the Software Products. Affymetrix shall have no obligation under this Agreement to correct any bugs, defects or errors in the Software Products or to otherwise support or maintain the Software Products. No warranty with respect to performance specifications of the Software is made. You acknowledge and agree that Affymetrix may modify or cease supporting the Software Products at any time with or without notice and the entire risk of using the Software Products is borne by you. AFFYMETRIX DISCLAIMS ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NON-INFRINGEMENT, RELATED TO THE SOFTWARE PRODUCTS, THEIR USE OR ANY INABILITY TO USE THEM, THE RESULTS OF THEIR USE AND THIS AGREEMENT.

7. Limitation on Damages. IN NO EVENT SHALL AFFYMETRIX OR ITS DISTRIBUTOR (IF APPLICABLE) BE LIABLE FOR ANY DAMAGES, WHETHER IN CONTRACT OR TORT (INCLUING NEGLIGENCE) OR UNDER ANY OTHER LEGAL THEORY, INCLUDING BUT NOT LIMITED TO DIRECT, CONSEQUENTIAL, SPECIAL, EXEMPLARY, INCIDENTAL AND INDIRECT DAMAGES, ARISING OUT OF OR IN CONNECTION WITH THIS AGREEMENT OR THE USE OR THE INABILITY TO USE THE SOFTWARE PRODUCTS OR THE RESULTS OF USING THE SOFTWARE PRODUCTS. THE LIABILITY OF AFFYMETRIX SHALL IN NO EVENT EXCEED FIFTY UNITED STATES DOLLARS (U.S. $50.00). YOU ACKNOWLEDGE AND AGREE THAT THE RISKS OF LOSS HEREUNDER ARE REFLECTED IN THE FACT THAT THE SOFTWARE PRODUCTS ARE LICENSED AT NO CHARGE TO YOU, AND THAT THESE TERMS WOULD HAVE BEEN DIFFERENT IF THERE HAD BEEN A DIFFERENT ALLOCATION OF RISK.

8. Compliance with law. You shall not receive, download, ship, transfer or otherwise export or re-export the Software Products or any underlying information or technology (or direct product thereof) except in full compliance with all United States and other applicable laws and regulations, including, but not limited to, the United States Export Administration Act or any other export laws, restrictions or regulations (collectively the "Export Laws"). In addition, if the Software Products are identified as export controlled items under the Export Laws, you represent and warrant that you are not a citizen of, or otherwise located within, an embargoed nation, and that you are not otherwise prohibited under the Export

Laws from receiving the Software Products. All rights to use the Software Products are granted on condition that such rights are forfeited if you fail to comply with the Export Laws or the terms of this Agreement.

9. Survival. Except for the License, except for the phrase "but only a limited right of use which is revocable in accordance with the terms of this Agreement" in Section 3(c), except for Section 4(b), except for Sections 5(a), (b) and (c), and except as otherwise expressly provided herein, the terms of the Agreement shall survive termination.

10. Governing Law. This Agreement shall be governed and interpreted in all respects by the laws of the United States of America and the State of California as such laws are applied to agreements entered into and to be performed entirely within California between California residents, without regard to conflicts of laws provisions thereof. The parties agree that the United Nations Convention on Contracts for the International Sale of Goods is specifically excluded from application to this Agreement. You irrevocably submit to the nonexclusive personal and subject matter jurisdiction of the state and federal courts located in Santa Clara County, California. Notwithstanding the foregoing, Affymetrix may apply to any court of competent jurisdiction for temporary or preliminary injunctive relief.

11. Severability. In the event that any provision of this Agreement shall be unenforceable or invalid under any applicable law or be so held by a court of competent jurisdiction, such unenforceability or invalidity shall not render this Agreement unenforceable, or invalid as a whole, and, in such event, any such provision shall be changed and interpreted so as to best accomplish the objectives of such unenforceable or intended provision within the limits of applicable law or applicable court decisions.

12. U.S. Government End Users. If you are an agency, department, or other entity of the United States government ("Government"), the use, duplication, reproduction, release, modification, disclosure or transfer of this product, or of any related documentation of any kind, including technical data, is restricted in accordance with Federal Acquisition Regulation ("FAR") 12.212 for civilian agencies and Defense Federal Acquisition Regulation Supplement ("DFARS") 227.7202 for military agencies. The Software Products are commercial computer software and commercial computer software documentation. The use of the Software Products by the Government is further restricted in accordance with the terms of this Agreement. The contractor/manufacturer is Affymetrix Inc., 3380 Central Expressway, Santa Clara, CA 95051.

13. Miscellaneous. (a) You agree that (i) a material breach of this Agreement adversely affecting Affymetrix's proprietary rights in the Software Products or (ii) a breach of confidentiality provisions in this Agreement, would cause irreparable injury to Affymetrix for which monetary damages would not be an adequate remedy and that the Affymetrix shall be entitled to equitable relief (including, without limitation, injunctions) in addition to any remedies it may have hereunder or at law. (b) The License and the Agreement are not assignable or transferable by you without the prior written consent of Affymetrix; any attempt to do so shall be void. However, Affymetrix may transfer or assign any of its rights or obligations under this Agreement without your consent. (c) No failure to exercise, and no delay in exercising, on the part of either you or Affymetrix, any privilege, any power or any rights hereunder will operate as a waiver thereof, nor will any single or partial exercise of any right or power hereunder preclude further exercise of any other right hereunder. (d) The prevailing party in any action to enforce this Agreement shall be entitled to recover costs and expenses, including, without limitation, attorneys' fees. (e) YOU ACKNOWLEDGE AND AGREE THAT THE WARRANTY DISCLAIMERS AND LIABILITY AND REMEDY LIMITATIONS IN THIS AGREEMENT ARE AMONG THE MATERIAL BARGAINED FOR BASES OF THIS AGREEMENT AND THAT THEY HAVE BEEN TAKEN INTO ACCOUNT AND REFLECTED IN DETERMINING THE CONSIDERATION TO BE GIVEN BY YOU IN CONNECTION WITH THE SOFTWARE PRODUCTS AND IN YOUR DECISION TO ENTER INTO THIS AGREEMENT. (f) This Agreement represents the entire agreement concerning the Software Products between you and Affymetrix, and it supersedes any prior proposal, representation, or understanding between you and Affymetrix. (g) Any waivers or amendments shall be effective only if made in writing, clearly understood by you and Affymetrix to be an amendment or waiver and signed by representatives authorized to bind you and Affymetrix respectively.