Vector Xpression™ 3

Speed Tutorial:

Part I. Importing Two-Channel Raw Data

InforMax®

Invitrogen™ life science software

Table of Contents

Important: Please Read

STOP: 	This tutorial assumes that you are familiar with the standard Windows user interface and basic Windows techniques, such as maximizing windows, selecting objects, zooming in and out on objects, switching between panes in a viewer window, etc. For more information about basic Windows operations, see Chapter 3 of the Vector Xpression™ 3 User's Manual. It also assumes that you are somewhat familiar with gene expression techniques and data. For more information, see the "Gene Expression Overview" section in Chapter 4 of the *Vector Xpression™ 3 User's Manual.* Before beginning this tutorial, complete the following actions if you have not yet performed them: 1. Install Vector Xpression 3. For more information about installing Vector Xpression 3, see the *Vector Xpression™3 Installation Guide* that can be accessed from the InforMax Web site: http://www.informaxinc.com/vnti/vntisuite/Installation_VXpression <date>.pdf 2. In Windows, click **Start > Settings > Control Panel**. Click **Display**. In the Display Properties dialog box, make sure that **Colors** is set to a minimum of **High Color (16 bit)**. (Lower settings will cause scatter plots shown in this tutorial to display black.)

GO: 	If you have completed Steps 1 and 2 listed above, proceed with this tutorial.

Introduction
Purpose

This Speed tutorial teaches you to import two-channel expression data, producing a raw data object in the Vector Xpression database. It is part of a series of three tutorials to teach you how to use Vector Xpression:

I. Importing Two-Channel Expression Raw Data
II. Adding Annotations
III. Creating a Script for Automating Data Normalization

Other tutorials are available from InforMax to teach you other methods of using Vector Xpression. Refer to the InforMax website for more information:

http://www.informaxinc.com/content.cfm?pageid=25

Reference

The Speed tutorial is based on the microarray data produced and analyzed by Callow MJ, Dudoit S, Gong EL, Speed TP, and Rubin EM (2000). Microarray expression profiling identifies genes with altered expression in HDL-deficient mice. Genome Res *10*:2022-2029.

Experiment Overview

This experiment surveys the effects of a knockout of the apoAI gene in mice on the expression of other genes in the liver. Livers from mice homozygous for the apoAI knock out on a C57Bl/6 background as well as control C57Bl/6 mice were obtained, total RNA extracted, and the individual samples from each liver reverse transcribed using Cy5 dUTP. The reference RNA was a pool of total RNA obtained from the livers of eight C57Bl/6 mice, reverse transcribed with Cy3 dUTP.

To download this paper in pdf format, click the following link:

http://www.genome.org/cgi/content/full/10/12/2022

Expression Data Sets

Because this tutorial uses the same data analyzed in the publication, you can download a file containing all of the original expression data from the author's Web page:

http://stat-www.berkeley.edu/users/terry/zarray/Html/apodata.html

For the purposes of this tutorial, however, you will use six separate raw data files created from the original data text file. You can retrieve them from this site:

http://stat-www.berkeley.edu/users/terry/zarray/Software/smacode.html.

Click on the **tar file** link by the **Data Files** section. Click **Open** in the File Download dialog box, unzip the files and save them on your hard drive. Open each file in Notepad, and save it as a text file, indicated by the .txt extension to each file name.

Additional Notes About the Data Sets

The data sets used in this tutorial are a subset of the original data and are distributed as part of the Speed group's R program SMA (*S*tatistical *M*icroarray *A*nalysis) package. This subset of the original data was chosen for two reasons: 1) the data from the author's website is only available as an R data frame which needs extensive manipulation for converting into a format suitable for Vector Xpression; 2) the full apoA1 data frame is a summarized version of the original data, only containing the background corrected fluorescence intensity values.

The relationships between the files used in this tutorial and the columns in the file containing the full apoAI data set available from the author's website are listed in Table 1.

Data File	Sample (Cy5) channel	Reference (Cy3) Channel
sample.c1.txt	c1R	c1G
sample.c2. txt	c2R	c2g
sample.c3. txt	c3R	c3G
sample.t1. txt	k1R	k1G
sample.t2. txt	k2R	k2G
sample.t3. txt	k3R	k3G

Table 1 Components of text files comprising the Speed data sets

IMPORTANT: 	You may want to use this tutorial in conjunction with the *Vector Xpression™ 3 User's Manual* for clarification of all functionality.

Creating a New Vector Xpression Database

Overview

Vector Xpression saves information such as raw data, chip designs and layouts as database "objects," which are then stored in a relational database. The Vector Xpression Database Explorer organizes these expression objects for easy retrieval and management. Additionally, it supports intuitive browsing of databases, drag and drop operations, and other functions typical of window-based database management.

From the Explorer window, you can perform the following operations:

- Open raw data objects, Expression Runs, Runs Projects and expression experiments
- Sort, edit and delete items
- Search the database
- Organize your data into convenient groups (subsets)
- Exchange database objects among databases and subsets
- Import and export data

You will use the Database Explorer to import the data.

Action

1. From the Windows Start button, select **Start > Programs > InforMax 2003 > Xpression Explorer** to open the Vector Xpression Database Explorer (Figure 1).

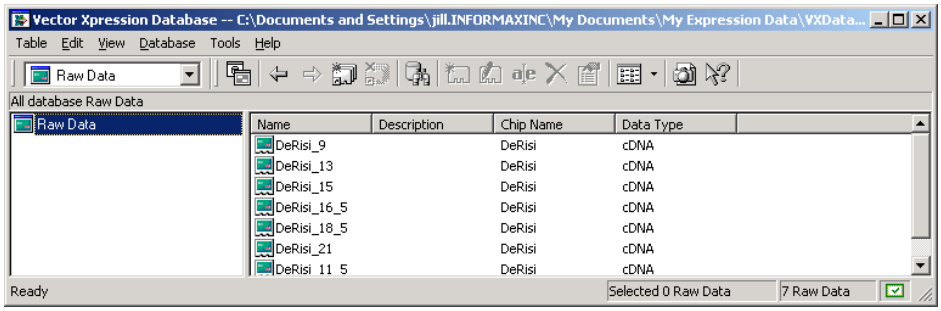

Figure 1 Opening Vector Xpression Database

2. From the Vector Xpression Database menu bar, select **Database > New Empty Database**.

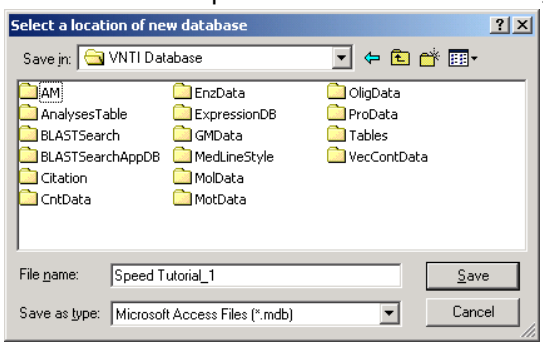

Figure 2 Select a Location of a New Database Dialog Box

3. In the Select a Location of New Database dialog box (Figure 2), name the new database file ***Speed_Tutorial_1***, navigate to a convenient drive and folder for saving the new file, and click **Save**.

Result

This returns you to the Vector Xpression Database viewer window displaying the new, blank database. Now you will load the first raw data set into the empty database.

Opening the Raw Data Source File

Overview

Using a raw data text file, you will create an import scheme and load the raw data into Vector Xpression using the Import Tool.

Vector Xpression's dual-purpose tool called Import creates the necessary import scheme for an expression data file and then uses that scheme to import the expression data file into Vector Xpression. This produces a Raw Data object in the Vector Xpression database.

Action

1. In the Vector Xpression Database Explorer, select **Tools > Import Expression Data**.

   The Select File(s) to Import dialog box displays (Figure 3).

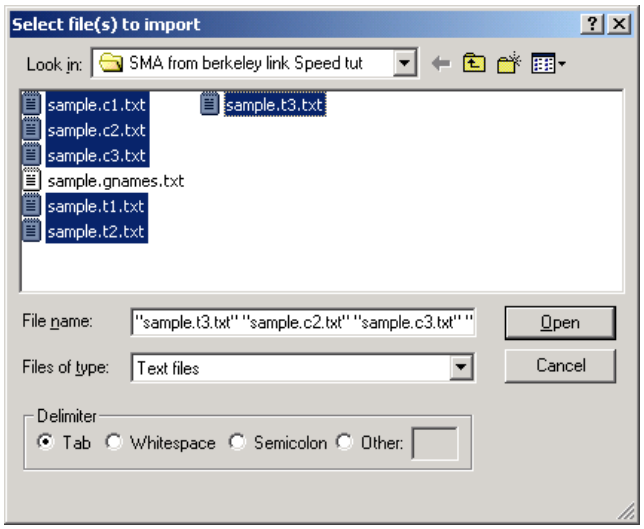

Figure 3 Select Expression Data File(s) to Import Dialog Box

   a. In the **Look in** list, navigate to the directory where you saved the source data files. (Do not click **Open** yet.)
   b. In the **Files of type** list, select **Text files**.
   c. CONTROL + CLICK the **Sample.c1, c2 and c3.txt** files and the **Sample.t1, t2 and t3.txt** files in the area below the **Look in** list.
   d. In the **Delimiter** area, accept **Tab**.
   e. Click **Open**.

2. This opens the Assign Schemes dialog box (Figure 4).

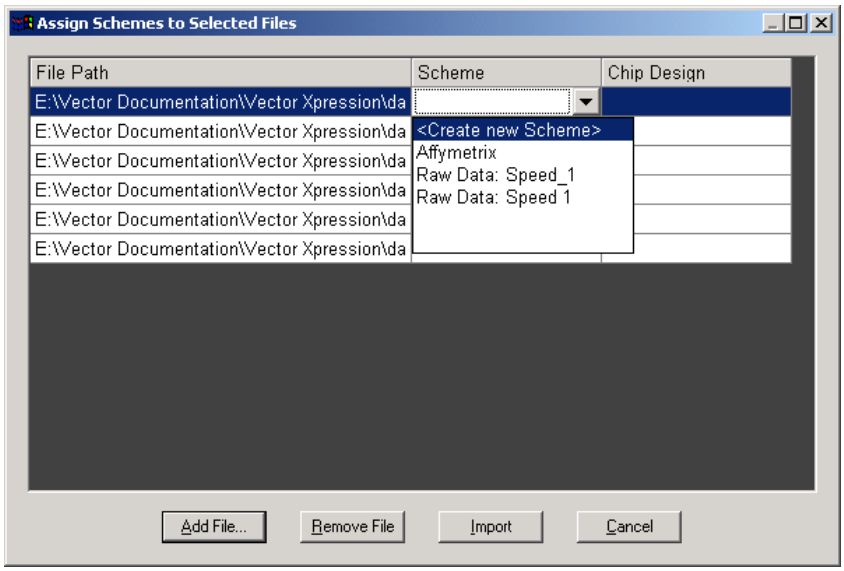

Figure 4 Assign Schemes to Selected Files dialog box

a. Click in the upper-most empty text box in the **Scheme** column.
b. On the drop-down list that appears, select **<Create new Scheme>** (Figure 5).

3. In the Import Data File Type dialog box that opens, select the **Raw Data** radio button and click **OK**. (Figure 6).

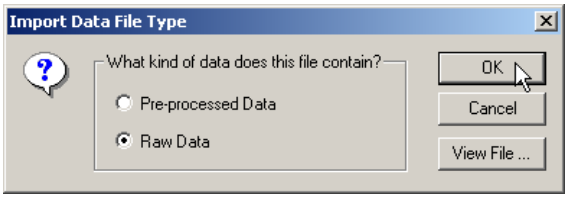

Figure 5 Selecting the data type to be imported

The Import window containing a spread-sheet view of the selected raw data source file opens (Figure 6). It is superimposed by the Header and Data dialog box.

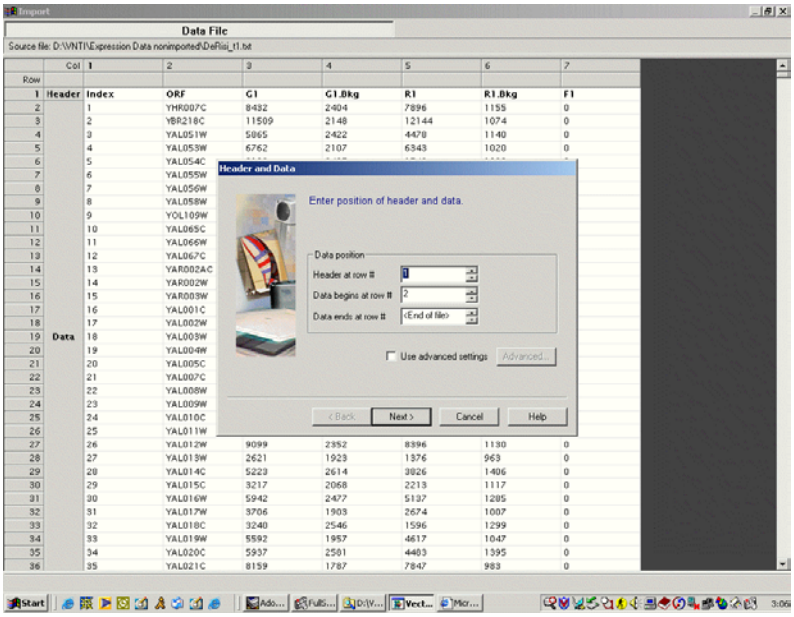

Figure 6 Import Dialog Box – Data from Source File

Result

You have successfully opened the raw data source file. Now you will view and define the raw data source file for import.

Viewing and Defining Raw Data in the Source File

Overview

In configuring an expression data file for import, the main Import dialog box opens displaying the source file. After your source file is loaded, browse the source data to familiarize yourself with it. This source file remains open for your referral as you step through each window using the Import Wizard.

The source file used to build the import scheme presents a complete dataset of one microarray experiment. Data presented in the source file are absolute signals, real measurements read from a chip containing a specific number of genes. One-channel data represents values read from one experiment. Two-channel data represents one experiment in which RNA from a sample was labeled during reverse transcription with the red-fluorescent dye Cy5, and was mixed with a sample labeled in parallel with the green-fluorescent dye Cy3.

This tutorial uses data from two-channel experiments.

Action

The first dialog box of the Import Wizard, the Header and Data dialog box, is superimposed on the Import window. To better review the source file, click on the blue title bar of the dialog box and drag it out of the way.

1. In the open source file, note the location of the Header row and Data start and end rows (Figure 7). The Header is the first row in the file and usually contains the column names. The data is contained in rows 2 through 6385.

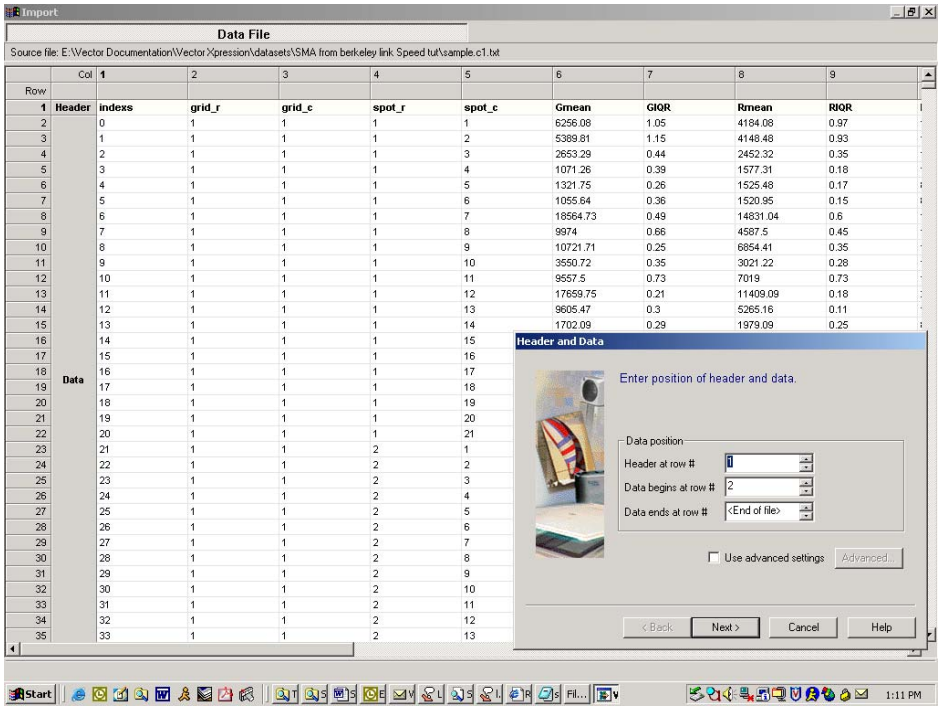

Figure 7 Open source file displaying experimental data

2. In the Header and Data dialog box (Figure 8), verify or enter the following:
   a. Enter the appropriate rows in the **Header** and **Data** text boxes as shown in Figure 9 (if they are not entered by default). *Selecting **End of file** for **Data end…** instead of specifying the row containing the end of the data allows flexibility when this scheme is used on future files.*

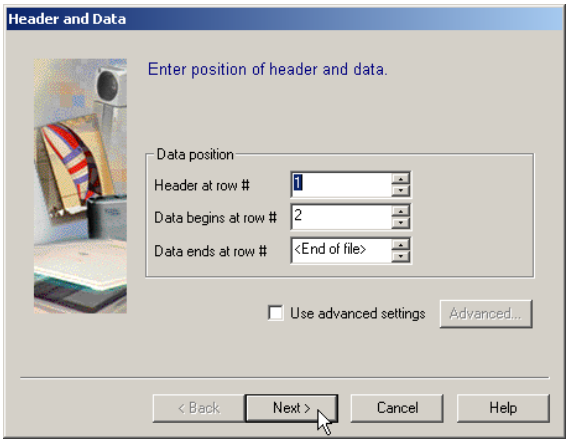

Figure 8 Header and Data dialog box

   b. Click **Next** to continue.
3. The Select Number of Channels dialog box opens (Figure 9):

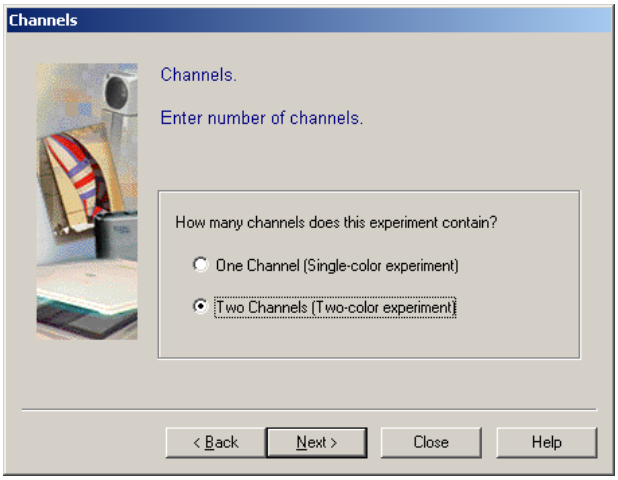

Figure 9 Channels Dialog Box

   a. Select the **Two Channels (Two-color experiment)** radio button.
   b. Click **Next** to continue.
4. The **Data** dialog box opens (Figure 10):

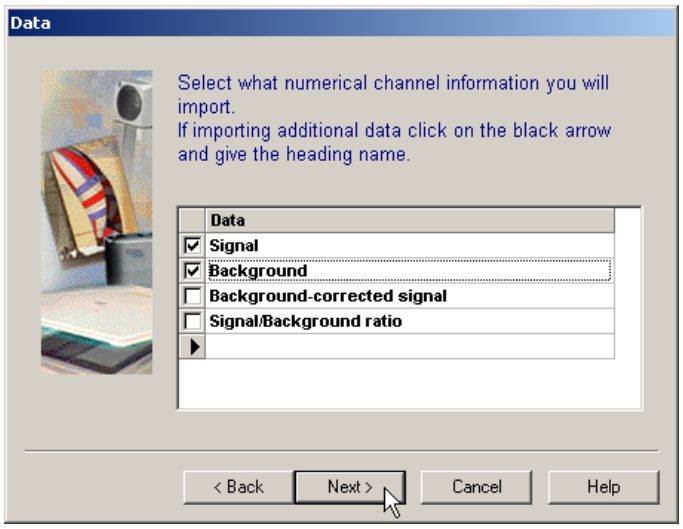

Figure 10 Data (Enter the data you will import) Dialog Box

c. Select the **Signal** and **Background** check boxes.

Signal and Background are the names for the relevant data columns to be used in Vector Xpression. In the source file (Figure 7 on page 7), note that columns 6 and 10 contain signal and background data for one channel, and columns 8 and 13 contain signal and background data for the other channel.

d. Click **Next** to continue.

5. The Assign Columns to Channel Data dialog box opens:

e. Enter the appropriate column numbers in the text boxes as shown in Figure 11. Verify the entries by reviewing the corresponding data in the source file.

To enter the information easily, click on each appropriate column header in the spreadsheet view and drag it to the respective text box. To enter figures manually, click twice in each cell, activating the text box, and type the correct number.

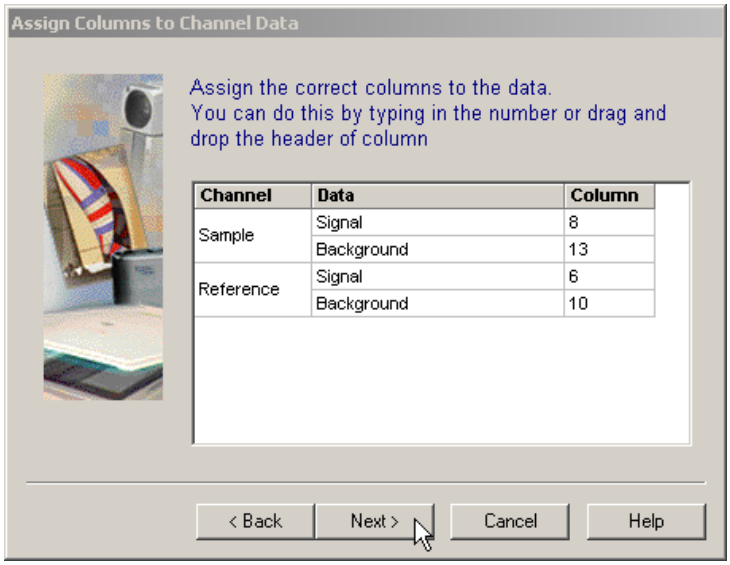

Figure 11. Assign Columns to Channel Data Dialog Box

f. Click **Next** to continue.

6. In the Additional Spot Data dialog box that opens, verify that the checkboxes are unchecked. Click **Next >** to continue.
7. In the second Additional Spot Data dialog box that opens, verify that the check boxes are unchecked. Click **Next >** to continue.
8. The Chip Design Information Source dialog box opens to begin the next phase of the import process.

Result

You have successfully viewed and defined the raw data source file. Now you will build the chip design and associate it with the raw data.

Building the Chip Design

Overview

To continue the import, you must build the chip design and associate it with the raw data. Chip designs link the spot locations of Raw Data objects to their gene names.

Action

1. The Chip Design Information Source dialog box allows you to identify the source file that provides information about the configuration of a given microarray chip. If the file were the file you already have open, you would select the radio button **This File**. In this case, however, select **Another file**, and click the Browse button (...).
2. In the Select file(s) to Import dialog box (Figure 12), navigate to the folder containing the chip design file to be loaded.

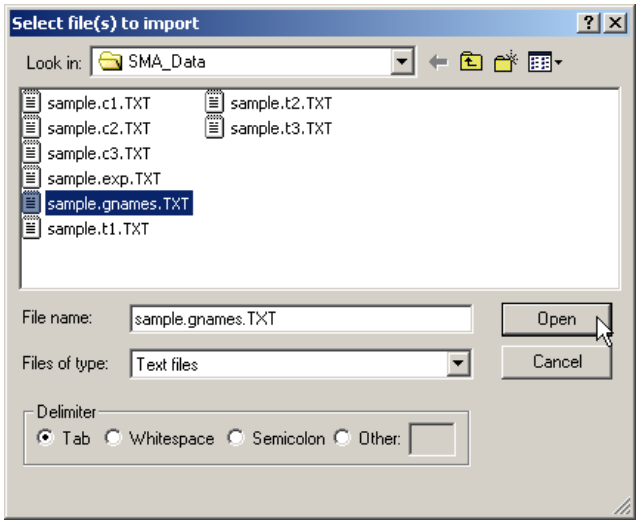

Figure 12 Select the Chip Design File for your expression data

a. Select the file **sample.gnames.TXT**, whose name now displays in the **File name** field.
b. Select **Text files** from the drop-down list in the **Files of type** text box.
c. From the **Delimiter** radio buttons, select **Tab**.
d. Click the **Open** button.

3. Back in the Chip Design Information Source dialog box, click **Next >** to continue.

4. The Coordinates dialog box is now superimposed on the source file.
   a. Select the **Spot Coordinates** radio button.
   b. Drag the column header named **grid_r** to the top text box in the **Coordinate** column.
   c. Drag the column headers in the spreadsheet named **grid_c**, **spot_r**, and **spot_c** to consecutive rows, as illustrated in the dialog box in Figure 13.

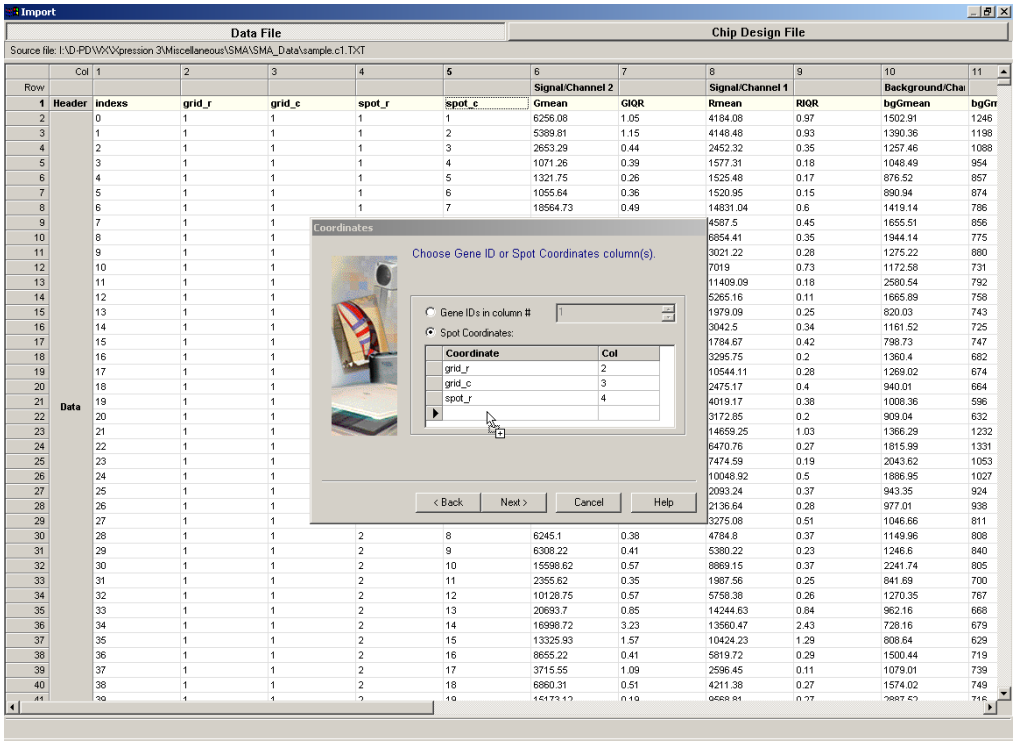

Figure 13 Chip Design Window

Still in the Coordinates dialog box, now you will change the terminology for your selections (Figure 14).

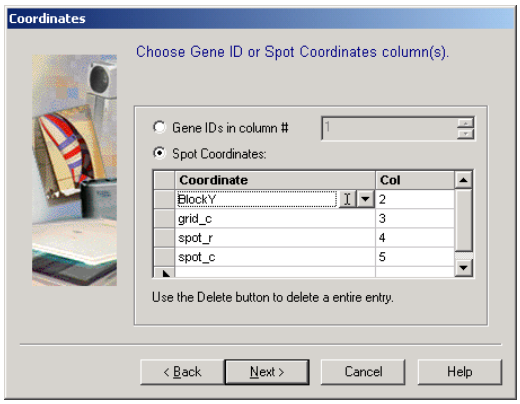

Figure 14 Coordinates Dialog Box

d. Click in the cell labeled **grid_r**.

e. Click the down arrow and select **BlockY** from the drop-down list.

f. Repeat the process for the **grid_c**, **spot_r**, and **spot_c** entries, replacing them with **BlockX**, **Y**, and **X** respectively.

d. Click **Next >** to continue.

At this point, the Chip Design File button in the Import window becomes acti, and the contents of the chip design file now display in the spreadsheet.

5. The Header and Data dialog box opens (Figure 15):

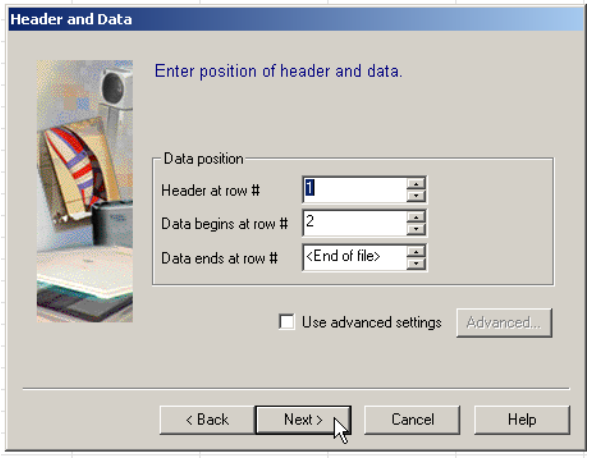

Figure 15 Header and Data dialog box

a. Enter the appropriate rows in the text boxes as shown in Figure 15 (if they are not entered by default).

b. Click **Next >** to continue.

7. Insure that the **Data File** button is still active in the Import window.

8. The Coordinates dialog box opens (Figure 16):

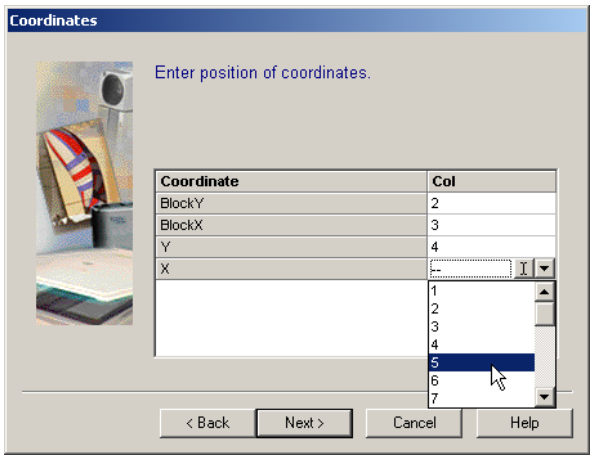

Figure 16. Designate column coordinates in the Coordinates dialog box

   a. Click in the text box in the **Col** column adjacent to **BlockY** coordinate, and in the drop-down list that appears, select **2**.
   b. Repeat the process, selecting **3**, **4**, and **5** for the **BlockX**, **Y**, and **X** rows, respectively, as illustrated in Figure 21.
   c. Click **Next** to continue.

9. In the Gene Name column dialog box (Figure 17),
   a. Enter ***8*** in the **Column #** text box.

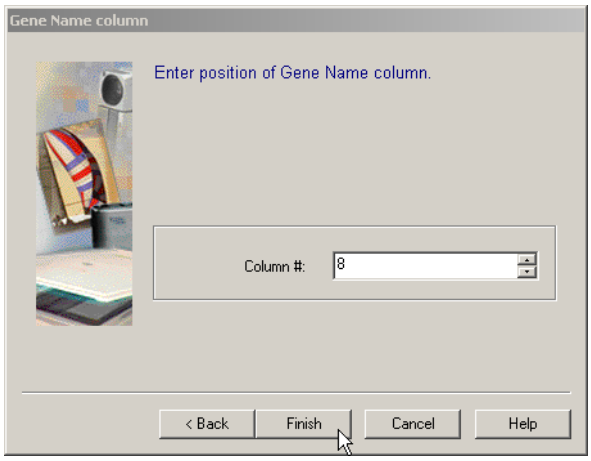

Figure 17 Gene Name dialog box

   b. Click **Finish**. In the dialog box that opens, enter the Scheme Name to continue.

At this point in the tutorial, you will import all of the text files necessary to complete the series of three Speed tutorials.

1. You are returned to the Assign Schemes to Selected Files dialog box (Figure 18):
   a. Click in the uppermost cell in the **Scheme** column, then SHIFT + CLICK on the lower-most cell. This selects all of the rows, and a drop-down text box displays in the lowest box.
   b. In the text box drop-down menu, select **Raw Data: Speed_1** import scheme.

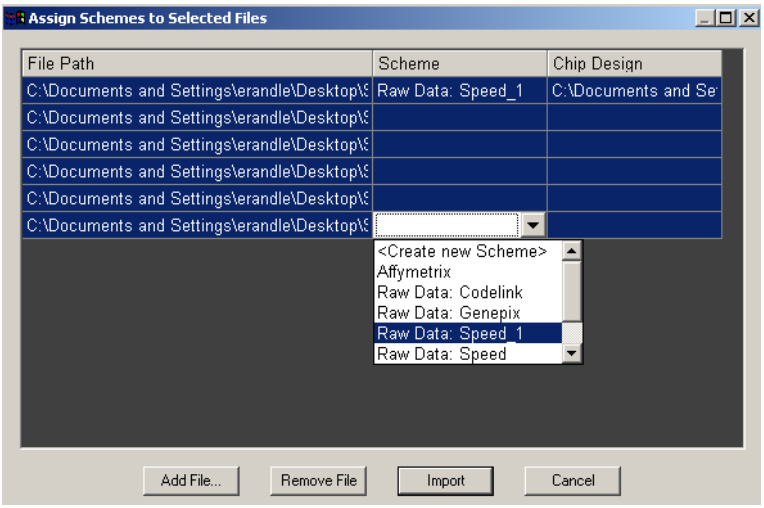

Figure 18 Assign Schemes dialog box, selecting the import scheme

c. Click the **Import** button. The Finalize Import dialog box opens.

Result

The Assign Schemes to Selected Files dialog box returns to view. The Chip Design cell in this dialog box now indicates that the chip design information is derived from the source file opened in the Import dialog box. It is assigned to all six files you are importing. Now you will process the data in the file for mapping into the Vector Xpression database.

Finalizing Import

Overview

The Finalize Import dialog box in which you will finalize import displays a summary of the file features for the database and allows you to change them, if necessary. It also allows you to name the column headers used in the spreadsheets of the Vector Xpression viewers.

This dialog box has three tabs, and it opens to the tab appropriate to the data type you are importing.

If necessary, left-click and drag the window borders or the dividers between the column headings to readjust the window size or column widths to see the column contents.

IMPORTANT:	In the Finalize Import dialog box cells, there are three methods you can use to enter or edit information for the specified data. Click in a blank cell or a cell with text you want to edit. One or more of three options may appear. • Click on the down arrow ( ) to reveal a drop-down list. • Click on the Browse button ( ) to open a navigation dialog box. • Click on the Text icon ( ) to activate the text box where you can enter or edit text.

2. In the Finalize Import dialog box, in the blank cell in the Chip column, select the browse button ( ) (Figure 19).

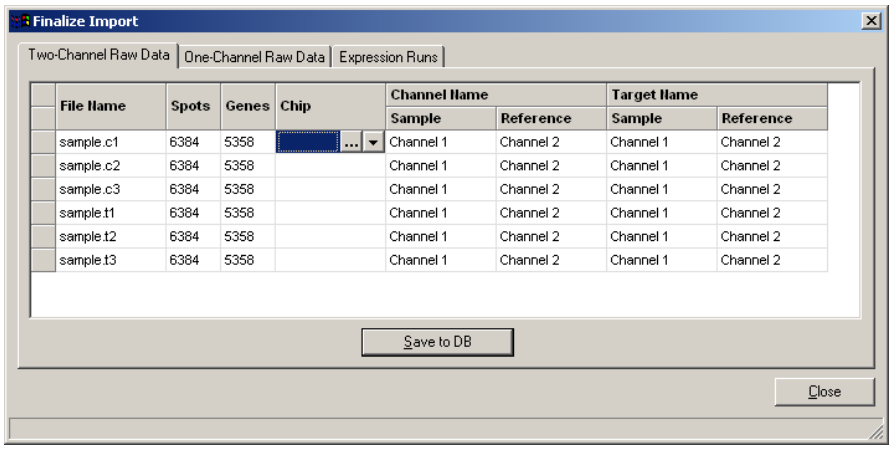

Figure 19 Finalize Import dialog box, assigning the chip

3. In the Create New Chip dialog box, name the new chip object ***Speed_1*** and click the **Create** button (Figure 20).

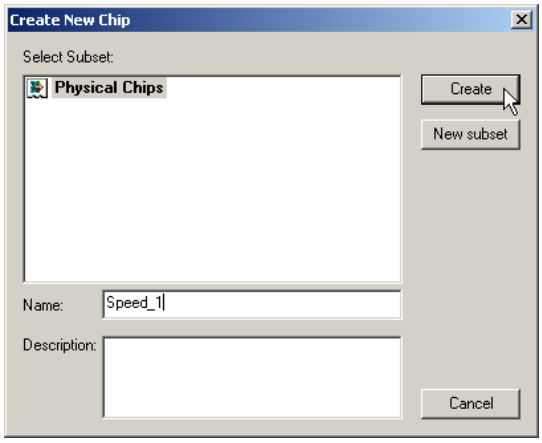

Figure 20 Create New Chip Dialog Box

Back in the Finalize Import dialog box, you can see in the **Chip** column the name of the new chip you've just created.

4. Use SHIFT + CLICK to select all the rows in the **Chip** column and choose **Speed_1** from the dropdown menu to assign this chip to all six files

Assigning the chip as you have just done is the only essential step required for this part of the import process. You can, however, edit other fields in the dialog box using the edit techniques previously described.

5. Click **Save to DB**.
6. In the Choose Subset dialog box, only the top-level Raw Data table (the root directory) is listed (Figure 21). Name a new subset in the Name text box ***Speed Tutorial*** and click **OK**. This is the subset where the data is to be stored.

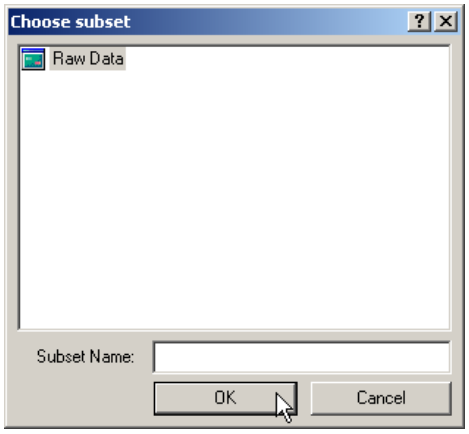

Figure 21 Select Subset Dialog Box

7. When prompted, select the **Yes** button to close the Import dialog box. This verifies that the data you just imported resides in the Vector Xpression Database.

   You are returned to the Vector Xpression Database Explorer (Figure 22). With the Raw Data table selected, you can see all the raw data objects you have just imported, with their chip name **Speed_1**, assigned to each.

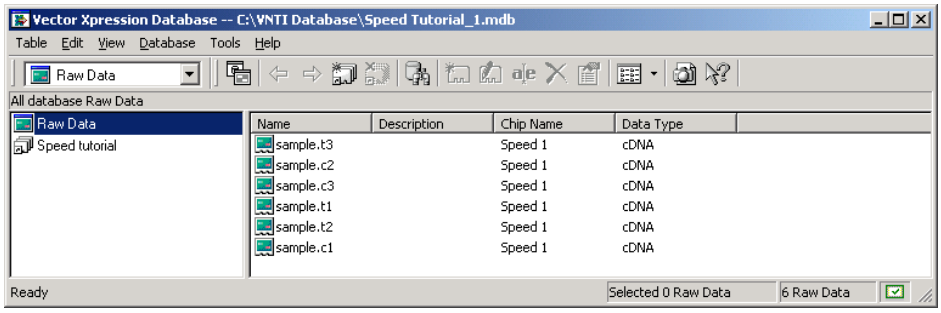

Figure 22 The Xpression Database Explorer displays imported files.

Result

You have successfully built the chip design, associated it with six files of raw data, and saved the chip to the Vector Xpression database.

Now you will continue to **Part II: Adding Annotations** (optional) and/or **Part III: Creating a Script for Automating Data Normalization**.