# Vector Xpression<sup>™</sup> 3

**Speed Tutorial:** 

Part I. Importing Two-Channel Raw Data



# Table of Contents

Important: Please Read	1
Creating a New Vector Xpression Database	3
Opening the Raw Data Source File	4
Viewing and Defining Raw Data in the Source File	7
Building the Chip Design	10
Finalizing Import	14

# 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 <sup>™</sup> 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 <i>Vector Xpression</i> ™ 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 <i>Vector Xpression</i> ™3 <i>Installation Guide</i> that can be accessed from the InforMax Web site:
	http://www.informaxinc.com/vnti/vntisuite/Installation_VXpression <date>.pdf</date>
	<ol> <li>In Windows, click Start &gt; Settings &gt; 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.)</li> </ol>



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

IntroductionThis Speed tutorial teaches you to import two-channel expression data,<br/>producing a raw data object in the Vector Xpression database. It is part of a<br/>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 <i>10</i> :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 C57BI/6 background as well as control C57BI/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 C57BI/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 <b>tar file</b> link by the <b>Data Files</b> section. Click <b>Open</b> 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 (Statistical *M*icroarray Analysis) 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  $\mathbb{M}$  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.



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

💽 Vector Xpression Database C	\Documents and !	Settings\jill.INFOR	MAXINC\My Docu	ments\My Express	sion Data\VXData	
Table Edit View Database Tools	Help					
📄 Raw Data 💽 🗌 🖷	4 ⇒ ఐ	) <b>q.</b>   to d	) aje 🗙 😭	⊞ • 🗿 🌾		
All database Raw Data						
🔜 Raw Data	Name	Description	Chip Name	Data Type		
	🔜 DeRisi_9		DeRisi	cDNA		
	🔜 DeRisi_13		DeRisi	cDNA		
	Risi_15		DeRisi	cDNA		
	DeRisi_16_5		DeRisi	cDNA		
	DeRisi_18_5		DeRisi	cDNA		
	DeRisi_21		DeRisi	cDNA		
	DeRisi 11 5		DeRisi	cDNA		-
Ready				Selected 0 Raw Data	7 Raw Data	

Figure 1 Opening Vector Xpression Database

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

Select a location of ne	ew database		? ×
Save in: 🔁 VNTI Da	tabase	- 🖬 📩 🗢 🔽	
AnalysesTable BLASTSearch BLASTSearch Citation Citation CntData	EnzData     ExpressionDB     GMData     MedLineStyle     MolData     MolData     MotData	CigData ProData Tables VecContData	
File <u>n</u> ame: Speed 1	utorial_1	<u>S</u> a	/e
Save as type: Microso	ft Access Files (*.mdb)	▼ Can	cel //

Figure 2 Select a Location of a New Database Dialog Box

 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).

Select file(s) to import	? ×
Look in: 🔁 SMA from berkeley link Speed tut 💽 🖛 🗈 📸 🏢	•
sample.c1.txt     sample.t3.txt       sample.c2.txt     sample.c3.txt       sample.gnames.txt     sample.t1.txt       sample.t2.txt     sample.t2.txt	
File name: ["sample.t3.txt" "sample.c2.txt" "sample.c3.txt" "	ben
Files of type: Text files Ca	ncel
Oelimiter	

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).

File Path	Scheme	Chip Design
E:\Vector Documentation\Vector Xpression\da		
E:\Vector Documentation\Vector Xpression\da	<create new="" scheme=""></create>	
::\Vector Documentation\Vector Xpression\da	Affymetrix	
:\Vector Documentation\Vector Xpression\da	Raw Data: Speed_1	
E:\Vector Documentation\Vector Xpression\da	····· - ···· - [ ·	
E:\Vector Documentation\Vector Xpression\da		

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

Import D	ata File Type	×
?	What kind of data does this file contain? Pre-processed Data Raw Data	OK Cancel View File

*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.

			Data File						
ource fi	e: D:\VNT	IVExpression D	ata nonimported\DeRis	_t1.txt					
	Col	1	2	3	4	S	6	7	
Row									
1	Header	Index	ORF	G1	G1.8kg	R1	R1.Bkg	F1	
2		1	YHR007C	8482	2404	7896	1155	0	
3		2	YBR218C	11509	2148	12144	1074	0	and the second sec
4		3	YAL051W	5065	2422	4470	1140	0	
5		4	YAL053W	6762	2107	6343	1020	0	
6		5	YAL054C			1.0.10			
7		6	YAL055W	eader and Date	•				
8		7	YAL056W						
9		8	YAL058W		Enter position (	of header and dat	a.		<ul> <li>State of the state of the state</li></ul>
10	10	9	YOL109W						
11		10	YAL065C	Contraction of the					
12		11	YAL066W						
13		12	YAL067C		- Data position				
14		13	YAR002AC	12	Header at som #	1			
15		14	YAR002W		rieduci di ionre	-			
16		15	YAR003W	-	Data begins at ro	ev # 2	-		
17		16	YAL001C	_		# Control Bas	-		
18		17	YAL002W		Data ends at row	# Frend or mess	-		
19	Data	18	YAL003W	-					
20		19	YAL004W				Leating Advance	. 1	
21		20	YAL005C			1 Use auvanue	settings	20d tor	
22		21	YAL007C						
23		22	YALOOSW						
24		23	YAL009W						a state of the state of the second
25		24	YAL010C		< Back	Next>	Cancel H	elp	and the second
26		25	YAL011W						
27		26	YAL012W	9099	2352	8396	1130	0	
28		27	YAL013W	2621	1923	1376	963	0	<ul> <li>A state of the sta</li></ul>
29		20	YAL014C	5223	2614	3026	1406	0	
30		29	YAL015C	3217	2068	2213	1117	0	
31		30	YAL016W	5942	2477	\$137	1205	0	
32		31	YAL017W	3706	1903	2674	1007	0	a second part and a second
9.9		32	YAL018C	3240	2546	1596	1299	0	
		33	YAL019W	5592	1957	4617	1047	0	A second state and second state and
34		34	YAL020C	5937	2501	4483	1395	0	and the family provided the second se
34		9-4							

Figure 6 Import Dialog Box – Data from Source File

#### Result

6).

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.

			Data	File								
ource fi	e: E:\Vect	or Documentat	ion/Vector Xpression	\datasets\SMA from I	berkeley link Speed to	ut\sample.c1.txt						
	Col	1	2	3	4	5	6	7	8	9		
Row												
1	Header	indexs	grid_r	grid_c	spot_r	spot_c	Gmean	GIQR	Rmean	RIQR		
2		0	1	1	1	1	6256.08	1.05	4184.08	0.97		
3		1	1	1	1	2	5389.81	1.15	4148.48	0.93		
4		2	1	1	1	3	2653.29	0.44	2452.32	0.35		
5		3	1	1	1	4	1071.26	0.39	1577.31	0.18		
6		4	1	1	1	5	1321.75	0.26	1525.48	0.17		
7		5	1	1	1	6	1055.64	0.36	1520.95	0.15		
8		6	1	1	1	7	18564.73	0.49	14831.04	0.6		
9		7	1	1	1	8	9974	0.66	4587.5	0.45		
10		8	1	1	1	9	10721.71	0.25	6854.41	0.35		
11		9	1	1	1	10	3550.72	0.35	3021.22	0.28		
12		10	1	1	1	11	9557.5	0.73	7019	0.73		
13		11	1	1	1	12	17659.75	0.21	11409.09	0.18		
14		12	1	1	1	13	9605.47	0.3	5265.16	0.11		
15		13	1	1	1	14	1702.09	0.29	1979.09	0.25		
16		14	1	1	1	15 H	eader and Data	Data				
17		15	1	1	1	16						
18		16	1	1	1	17	CONSTRAINTS OF	Enter position of	f handor and data			
19	Data	17	1	1	1	18		Enter position o	i fieader and uata.			
20		18	1	1	1	19						
21		19	1	1	1	20						
22		20	1	1	1	21		D 1 1				
23		21	1	1	2	1		- Data position				
24		22	1	1	2	2		Header at row #	0 -	-		
25		23	1	1	2	3		Databasia		-		
26		24	1	1	2	4		Data begins at rov	/#   <u>-</u>	-		
27		25	1	1	2	5		Data ends at row !	# <end file="" of=""></end>	-		
28		26	1	1	2	6	-					
29		27	1	1	2	7						
30		28	1	1	2	8			Use advanced set	tings Advanced	d	
31		29	1	1	2	9					-	
32		30	1	1	2	10						
33		31	1	1	2	11						
34		32	1	1	2	12		Contract C	N-1 0			
35		33	1	1	2	13		< васк	Next> Lar	icei He	яр	
						_						

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.

Header and Data	Enter position of header and data.
	Data position       Header at row #     Image: Comparison of the state of
	Use advanced settings Advanced
	K Back Next > Cancel Help

Figure 8 Header and Data dialog box

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

Channels	Channels. Enter number of channels.
	How many channels does this experiment contain?  C One Channel (Single-color experiment)  (Two Channels (Two-color experiment)
	< <u>B</u> ack <u>N</u> ext > Close Help

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):

Data	Sel imp If in and	lect what numerical channel information you will bort. nporting additional data click on the black arrow d give the heading name.
		Data
100 M		Signal
		Background
		Background-corrected signal
		Signal/Background ratio
		< Back Next > Cancel Help

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.

Assign Columns to	Channel Data Assign the You can do drop the hes	correct columns to the this by typing in the n ader of column	data. umber or drag and
	Channel	Data	Column
	Cavaula	Signal	8
	Sample	Background	13
	Pafaranaa	Signal	6
	Reference	Background	10
~			
	< Back	Next > Ca	ancel Help

Figure 11. Assign Columns to Channel Data Dialog Box

- f. Click **Next** to continue.
- In the Additional Spot Data dialog box that opens, verify that the checkboxes are unchecked. Click Next > to continue.
- 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.

Select file(s) to import	<u>? ×</u>
Look in: 🔁 SMA_Data 💽 🖛 🗈 📸 🖽 -	
sample.c1.TXT       sample.t2.TXT         sample.c2.TXT       sample.t3.TXT         sample.c3.TXT       sample.c3.TXT         sample.exp.TXT       sample.c1.TXT         sample.exp.TXT       sample.c1.TXT         sample.t1.TXT       sample.t1.TXT	
File name: sample.gnames.TXT Open	
Files of type: Text files Cance	
Delimiter © Tab © Whitespace © Semicolon © Other:	
	11.

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.

👷 🖥 Impo	rt											_ 8	×
			D	ata File					Chip Design F	ile			i
Source fi	le: I:\D-PD	WXXxpression	3\Miscellaneous\SM	IA\SMA_Data\sampl	le.c1.TXT								Π
	Col	1	2	3	4	5	6	7	8	9	10	11	
Row							Signal/Channe	12	Signal/Channel 1		Background/Cha	-	=
1	Header	indexs	grid_r	grid_c	spot_r	spot_c	Grnean	GIQR	Rmean	RIQR	bgGmean	bgGrr	
2		0	1	1	1	1	6256.08	1.05	4184.08	0.97	1502.91	1246	
3		1	1	1	1	2	5389.81	1.15	4148.48	0.93	1390.36	1198	
4		2	1	1	1	3	2653.29	0.44	2452.32	0.35	1257.46	1088	
5		3	1	1	1	4	1071.26	0.39	1577.31	0.18	1048.49	954	
6		4	1	1	1	5	1321.75	0.26	1525.48	0.17	876.52	857	
7		5	1	1	1	6	1055.64	0.36	1520.95	0.15	890.94	874	
8		6	1	1	1	7	18564.73	0.49	14831.04	0.6	1419.14	786	
9		7	1	1 6	ordinates				4587.5	0.45	1655.51	856	
10		8	1	1	oramaces				6854.41	0.35	1944.14	775	
11		9	1	1		Choose Gene ID (	or Spot Coordinates	column(s).	3021.22	0.28	1275.22	880	
12		10	1	1					7019	0.73	1172.58	731	
13		11	1	1					11409.09	0.18	2580.54	792	
14		12	1	1					5265.16	0.11	1665.89	758	
15		13	1	1		C Gene IDs in co	lumn # 1		1979.09	0.25	820.03	743	
16		14	1	1		C Seet Coordinal	, 		3042.5	0.34	1161.52	725	
17		15	1	1		se opor coordinal	cs.		1784.67	0.42	798.73	747	
18		16	1	1		Coordinate		Col	3295.75	0.2	1360.4	682	
19		17	1	1		grid_r		2	10544.11	0.28	1269.02	674	
20		18	1	1		grid_c		3	2475.17	0.4	940.01	664	
21	Bata	19	1	1		spot_r		1	4019.17	0.38	1008.36	596	
22	Data	20	1	1					3172.85	0.2	909.04	632	
23		21	1	1			1		14659.25	1.03	1366.29	1232	
24		22	1	1		_			6470.76	0.27	1815.99	1331	
25		23	1	1					7474.59	0.19	2043.62	1053	
26		24	1	1					10048.92	0.5	1886.95	1027	
27		25	1	1					2093.24	0.37	943.35	924	
28		26	1	1		< Back	Next > Cance	Help	2136.64	0.28	977.01	938	
29		27	1	1					3275.08	0.51	1046.66	811	
30		28	1	1	2	8	6245.1	0.38	4784.8	0.37	1149.96	808	
31		29	1	1	2	9	6308.22	0.41	5380.22	0.23	1246.6	840	
32		30	1	1	2	10	15598.62	0.57	8869.15	0.37	2241.74	805	
33		31	1	1	2	11	2355.62	0.35	1987.56	0.25	841.69	700	
34		32	1	1	2	12	10128.75	0.57	5758.38	0.26	1270.35	767	
35		33	1	1	2	13	20693.7	0.85	14244.63	0.84	962.16	668	
36		34	1	1	2	14	16998.72	3.23	13560.47	2.43	728.16	679	
37		35	1	1	2	15	13325.93	1.57	10424.23	1.29	808.64	629	
38		36	1	1	2	16	8655.22	0.41	5819.72	0.29	1500.44	719	
39		37	1	1	2	17	3715.55	1.09	2596.45	0.11	1079.01	739	
40		38	1	1	2	18	6860.31	0.51	4211.38	0.27	1574.02	749	
.1 .11		20	4	4	2	10	1517310	0.10	Q568 91	0.27	2887 52	716	*
•													

Figure 13 Chip Design Window

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

Coordinates			
	Choose Gene ID or Spot Coord	inates column(s).	
	Coordinate BlockY grid_c spot_r spot_c Use the Delete button to delete a e	Col A I V 2 3 4 5 V entire entry.	
	< <u>B</u> ack <u>N</u> ext >	Cancel Help	

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):

Header and Data	
	Enter position of header and data.
	Data position Header at row # Data begins at row # Data ends at row # CEnd of file>
	Use advanced settings Advanced
	K K Kext > Cancel Help

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):

Coordinates		
	Enter position of coordinates.	
	Coordinate	Col
	BlockY	2
And and a set	BlockX	3
	Y	4
	x	- I -
		1
		2
		3
	1	4
	< Back Next > Can	cel Help

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 <b>Column #</b> text box.
Gene Name colum	
	Enter position of Gene Name column.
	Column #: 3
	Kack Finish Cancel Help

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.

The Death	Oshama	Ohio Destina
ile Path	Scheme	Chip Design
:\Documents and Settings\erandle\Desktop\{	Raw Data: Speed_1	C:\Documents and Se
NDocuments and Settings\erandle\Desktop\		
NDocuments and Settings\erandle\Desktop\?		
:\Documents and Settings\erandle\Desktop\{		
:\Documents and Settings\erandle\Desktop\{		
NDocuments and Settings\erandle\Desktop\	•	
	<create new="" scheme=""> Affymetrix Raw Data: Codelink Raw Data: Genepix Raw Data: Speed_1 Raw Data: Speed</create>	•

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.

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

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

File Name	la llama Snote		Chin	Channel Nar	Channel Name		ə
	Spors	ocnes	CIIIP	Sample	Reference	Sample	Reference
sample.c1	6384	5358	🔻	Channel 1	Channel 2	Channel 1	Channel 2
sample.c2	6384	5358		Channel 1	Channel 2	Channel 1	Channel 2
sample.c3	6384	5358		Channel 1	Channel 2	Channel 1	Channel 2
sample.t1	6384	5358		Channel 1	Channel 2	Channel 1	Channel 2
sample.t2	6384	5358		Channel 1	Channel 2	Channel 1	Channel 2
sample.t3	6384	5358		Channel 1	Channel 2	Channel 1	Channel 2
				<u>S</u> ave to DB			

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).

Create New Chip	×
Select Subset:	
Physical Chips	Create
	New subset
Name: Speed_1	
Description:	
	Cancel

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.
- 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.

Choose subset		<u>? ×</u>
Raw Data		
Subset Name:	I	
	ок 💦	Cancel

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.

🚰 Vector Xpression Database C:\VNTI Database\Speed Tutorial_1.mdb							
Table <u>E</u> dit <u>V</u> iew <u>D</u> atabase Tools	<u>H</u> elp						
] 🔄 Raw Data 🔄   🔚 ( ⇔ ⇒ 🗊 🕼 ) 🖓   🖓 1 🗔 🖄 👘 🗡 🖆   ☶ - 🎒 💖							
All database Raw Data			_				
🔜 Raw Data	Name	Description	Chip Name	Data Type			
🗊 Speed tutorial	🔜 sample.t3		Speed 1	cDNA			
	🔜 sample.c2		Speed 1	cDNA			
	🔜 sample.c3		Speed 1	cDNA			
	🔜 sample.t1		Speed 1	cDNA			
	🔜 sample.t2		Speed 1	cDNA			
	🔜 sample.c1		Speed 1	cDNA			
Ready				Selected 0 Raw Data	6 Raw Data		

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**.