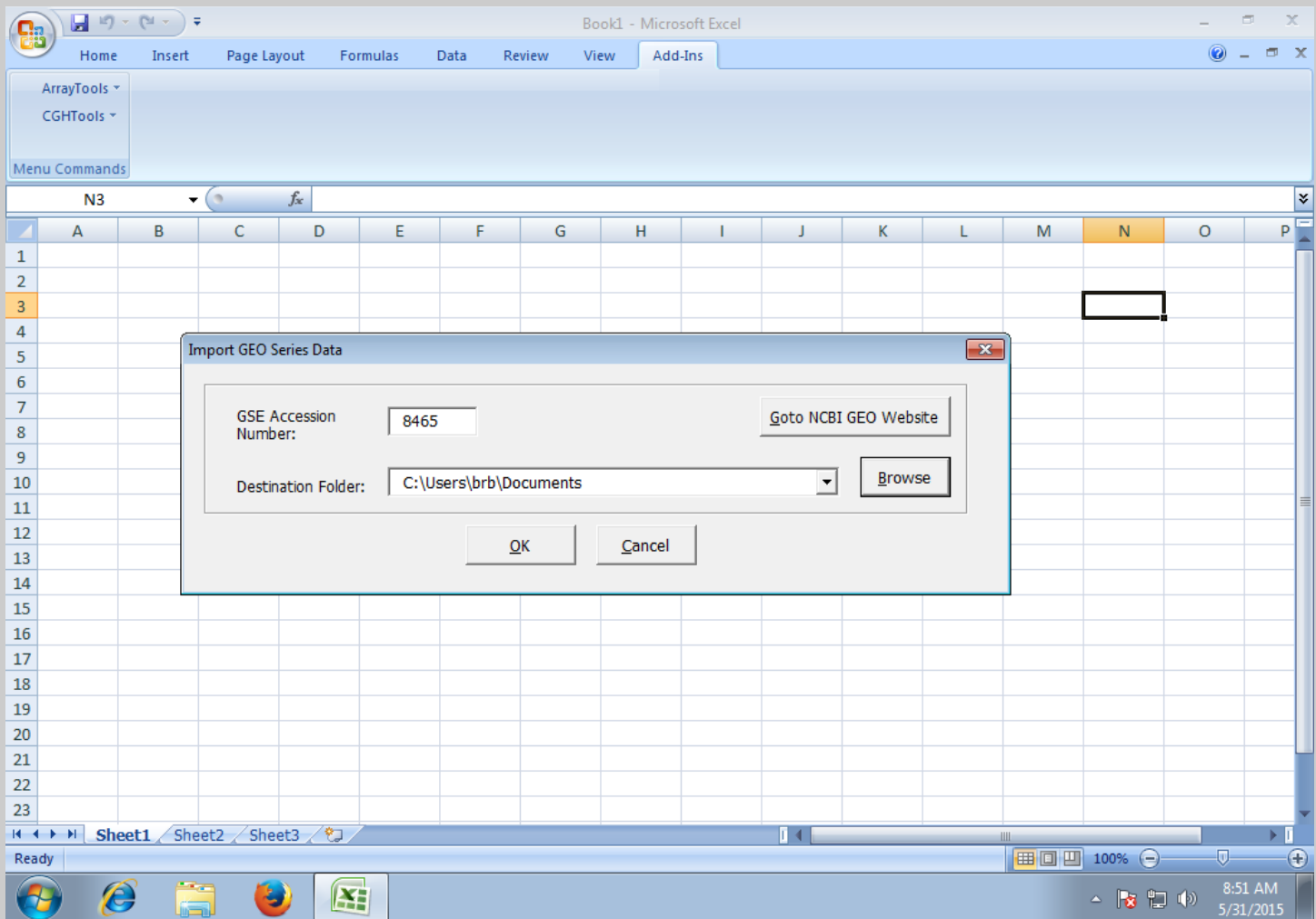
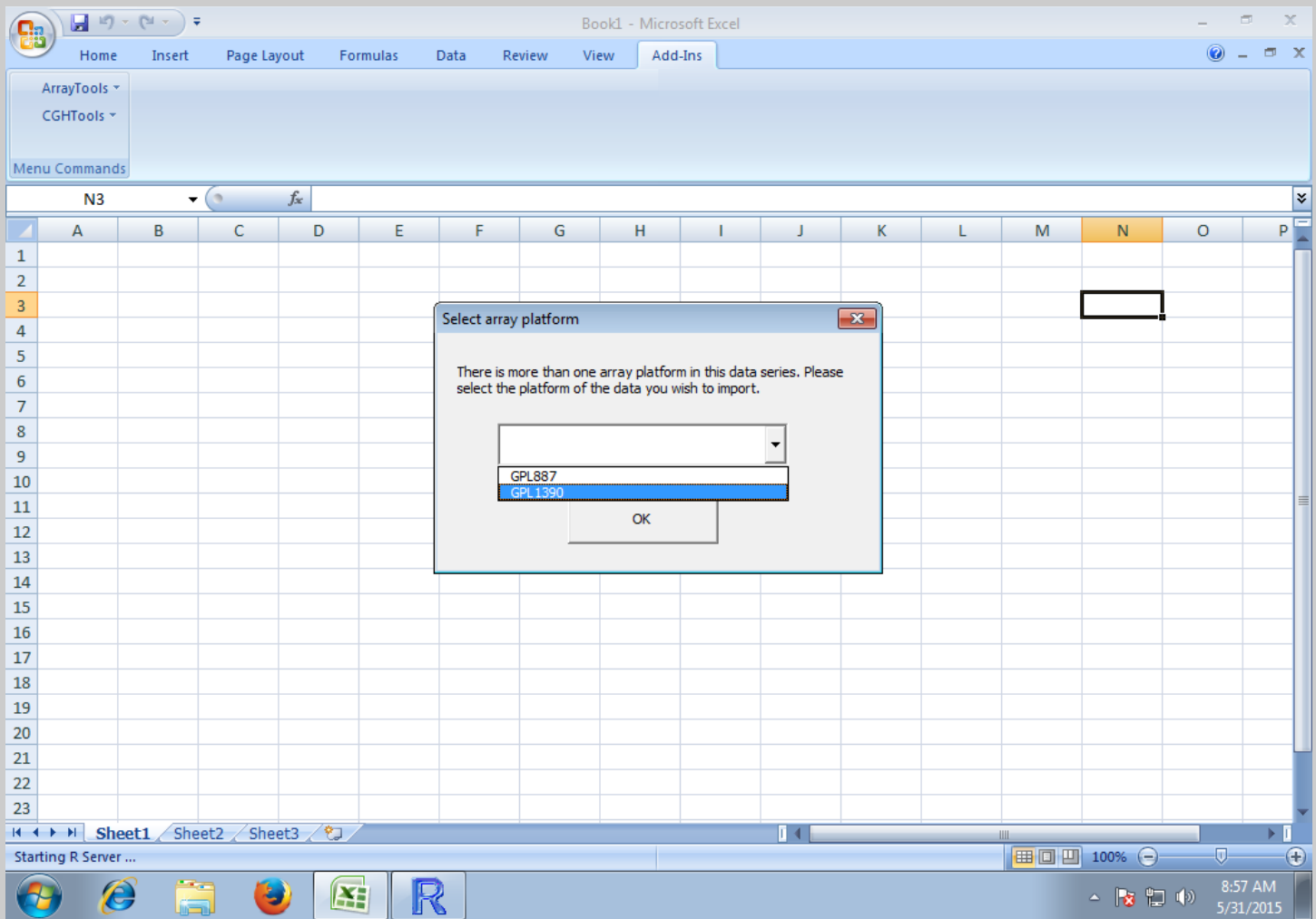


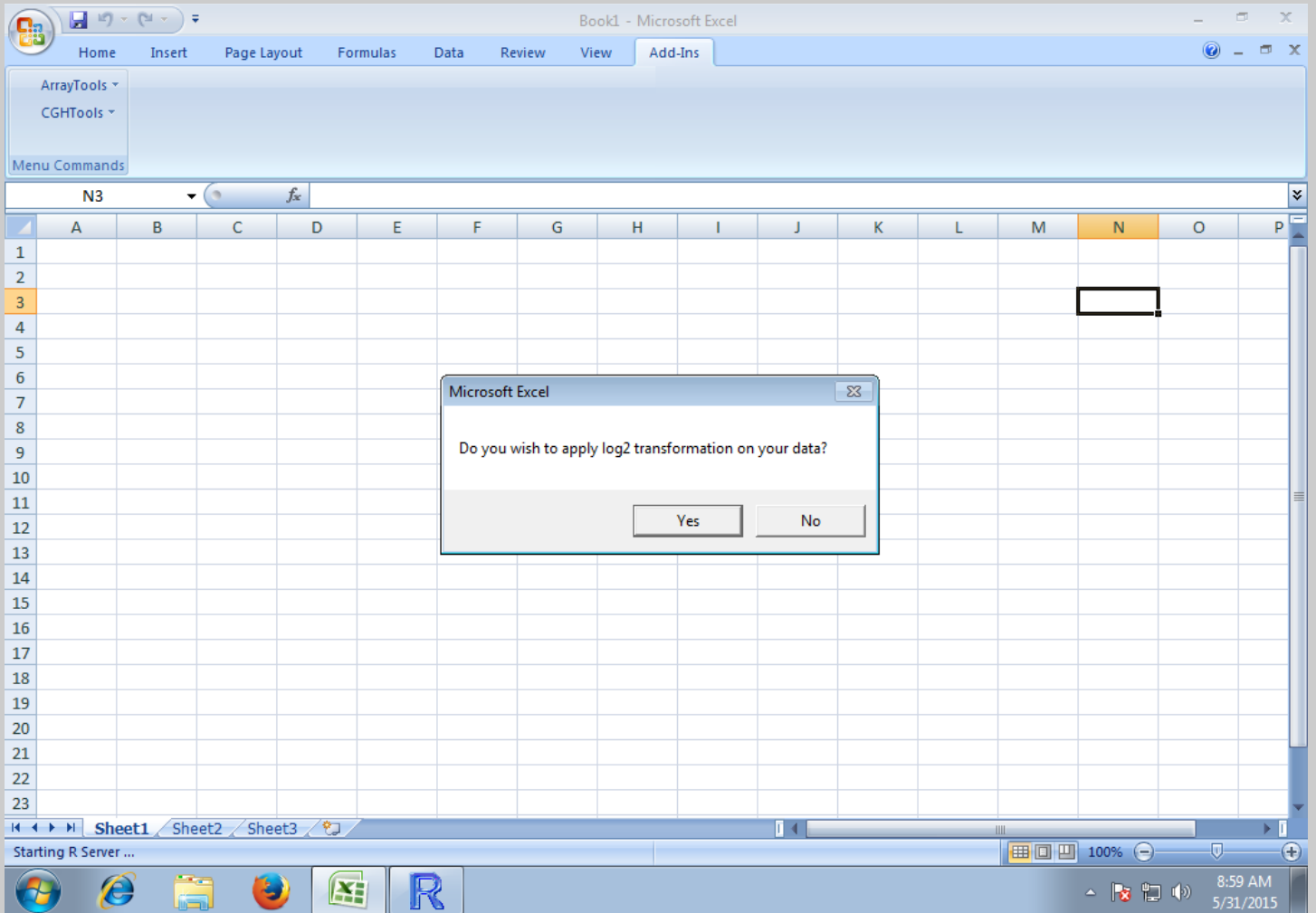
## Importing GSE series is a new feature in BRB-ArrayTools v4.5.0

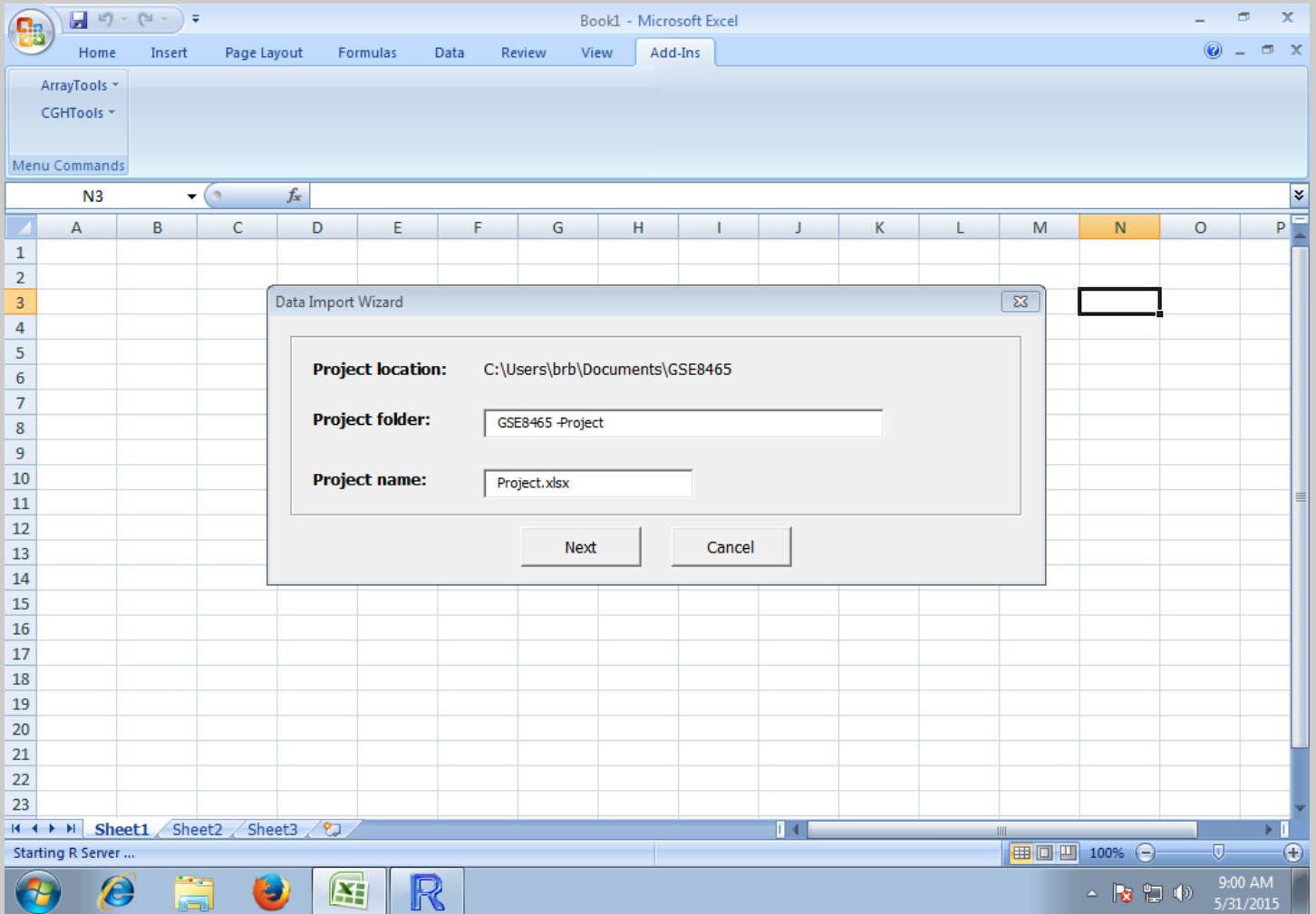
In this example, I pick up one GSE series contains more than one GPL.



If multiple GPLs was found, it will ask for the GPL number for the series belonging to that one.







The available gene identifiers depend on <GeneID.txt> file. In this case, five IDs were used.

**Data Import Wizard**

**Please specify the location of your gene identifiers:**

☐ The identifiers are stored alongside the expression data. ☒ The identifiers are stored in a separate file.

**Please select your Gene identifiers file:**

File: C:\Users\brb\Documents\GSE8465\GeneID.txt  Header line #: 1

**Please select the available gene identifiers:**

Unique ID (Well or Spot ID, etc.): Col 1: ID EntrezId:

Clone ID (IMAGE or ATCC ID, etc.):  Gene Name, Title or Description: Col 9: DESCRIPTION

UniGene Cluster ID: Col 10: Cluster ID GenBank Accession: Col 11: GB\_ACC

Gene Symbol: Col 7: GENE\_NAME Map Location:

Ensembl ID:  microRNA ID:

☐ Annotate the project with these gene ids, instead of using the data from SOURCE database. Organisms:

Since the <GeneID.txt> contains Ensembl column, we manually add it to the 'Ensembl ID' entry.

**Data Import Wizard**

Please specify the location of your gene identifiers:

☐ The identifiers are stored alongside the expression data. ☒ The identifiers are stored in a separate file.

Please select your Gene identifiers file:

File: C:\Users\brb\Documents\GSE8465\GeneID.txt  Header line #: 1

Please select the available gene identifiers:

Unique ID (Well or Spot ID, etc.): Col 1: ID EntrezId:

Clone ID (IMAGE or ATCC ID, etc.):  Gene Name ,Title or Description: Col 9: DESCRIPTION

UniGene Cluster ID: Col 10: Cluster ID GenBank Accession: Col 11: GB\_ACC

Gene Symbol: Col 7: GENE\_NAME Map Location:

Ensembl ID: Col 12: Ensembl Gene ID microRNA ID:

☐ Annotate the project with these gene ids, instead of using the data from SOURCE database. Organisms:

Project - Microsoft Excel

Home Insert Page Layout

ArrayTools  
CGHTools

Menu Commands

B2

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

A B C

M N O P

Experiment descriptors

Ready

9:03 AM  
5/31/2015

Refilter, normalize and subset the data

1. Spot filters 2. Normalization 3. Gene filters

Background adjustment is performed before the intensity filtering and the averaging of replicate spots is done on filtered data.

☐ Apply background adjustment.

☐ Intensity Filter:

☐ EXCLUDE the spot if BOTH intensities are below the minimum.

☐ EXCLUDE the spot if AT LEAST ONE of the two intensities is below the minimum.

☒ EXCLUDE the spot if BOTH intensities are below the minimum. If only ONE intensity is below the minimum, increase it to the minimum.

Red minimum: 100

Green minimum: 100

☐ Spot Flag Filter:

EXCLUDE the spot if the Spot Flag contains:

☒ Numeric values outside the range:

0 to 0

☐ Any of the following values:

List of values, separated by commas:

☐ Spot Size Filter:

EXCLUDE the spot if the Spot Size is less than: 10

☐ Average the replicate spots within an array.

☐ Use a common reference design. Change current setting

OK Cancel Reset Help

Hit the OK button to continue.

Project - Microsoft Excel

HomeInsertPage LayoutFormulasDataReviewViewAdd-Ins

ArrayTools  
CGHTools  
Menu Commands

B2fx

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	Display the data	GSM2100 26	GSM2100 63	Missing	P-Value	Rank	Variance	Num 1.5- Fold	Filter					
1	UniqueID													
7	6	-1.98736	-0.01342	0				2	TRUE					
17	16	-0.63227	-2.06839	0				2	TRUE					
19	18	-1.84377	0.854291	0				2	TRUE					
33	32	0.070188	-2.23133											
38	37	-6.0917	-3.2003											
51	50	1.601787	-0.37551											
60	59	-3.32617	-1.81237											
91	90	1.890153	-2.23133											
115	114	-4.0917	-1.38857											
125	124	-1.26388	1.18175											
143	142	1.109934	-0.81237	0				2	TRUE					
168	167	2.176023	-3.23133	0				2	TRUE					
173	172	-1.02131	-3.94753	0				2	TRUE					
185	184	-0.56188	-2.14018	0				2	TRUE					
187	186	-0.46526	-2.77761	0				2	TRUE					
192	191	-0.55565	-2.36257	0				2	TRUE					
209	208	0.066152	1.655564	0				2	TRUE					
224	223	-0.69253	-3.2003	0				2	TRUE					
237	236	-1.72713	-0.04064	0				2	TRUE					
270	269	1.932886	3.671377	0				2	TRUE					

Note

There are 3192 genes which pass the filtering and subsetting criteria.

OK

Experiment descripC:\Program Files\R\R-3.2.0\bin\x64\R.exeSheet3100%9:04 AM5/31/2015



Project - Microsoft Excel

Home Insert Page Layout Formulas Data Review View Add-Ins

ArrayTools  
CGHTools  
Menu Commands

B2

Proceed to Annotation?

You can proceed to annotate your data now. This might take a few minutes or longer. Please select one of the following options:

- ☐ Annotate Affymetrix data with chip-specific Bioconductor packages
- ☒ Annotate data with Bioconductor packages
- ☐ Annotate data through SOURCE
- ☐ Import annotation from an existing project with the same array type
- ☐ Annotate data later through 'Utilities' menu

OK Cancel

	A	B							
1	UniqueID	GSM21126							
7	6	-1.987							
17	16	-0.632							
19	18	-1.843							
33	32	0.0701							
38	37	-6.09							
51	50	1.6017							
60	59	-3.326							
91	90	1.8901							
115	114	-4.09							
125	124	-1.263							
143	142	1.1099							
168	167	2.1760							
173	172	-1.021							
185	184	-0.561							
187	186	-0.465							
192	191	-0.55565	-2.36257	0				2	TRUE
209	208	0.066152	1.655564	0				2	TRUE
224	223	-0.69253	-3.2003	0				2	TRUE
237	236	-1.72713	-0.04064	0				2	TRUE
270	269	1.932886	3.671377	0				2	TRUE
284	283	1.07447	0.335335	0				2	TRUE

Experiment descriptors Filtered log ratio Gene identifiers Sheet2 Sheet3

Ready

100%

9:05 AM 5/31/2015

The screenshot shows a Microsoft Excel window titled 'Project - Microsoft Excel'. The 'Add-Ins' tab is active, showing 'ArrayTools' and 'CGHTools' under 'Menu Commands'. The formula bar shows '= -0.61921203136444'. The active cell is B2, containing the text 'Display the data'. Below this, a table of data is visible, with columns A through N. The table has a header row with the following labels: UniqueID, GSM2100, GSM2100, Missing, P-Value, Rank, Variance, Num 1.5-Fold, and Filter. The data rows show values for these columns, with some cells highlighted in yellow. A dialog box titled 'Gene Annotation Engine' is open, displaying the text: 'Annotation is conducted using packages available at the Bioconductor website (www.bioconductor.org)'. It has two dropdown menus: 'Use gene identifier as lookup key:' and 'Organism:'. The 'Use gene identifier as lookup key:' dropdown is currently set to 'UGCluster', and the 'Organism:' dropdown is also set to 'UGCluster'. The 'Organism:' dropdown menu is open, showing options: 'UGCluster', 'Symbol', 'Accession', and 'Ensembl'. The Excel status bar at the bottom shows 'Ready', 'Experiment descriptors', 'Filtered log ratio', 'Gene identifiers', 'Sheet2', 'Sheet3', and a zoom level of 100%. The system tray at the bottom right shows the time as 9:06 AM on 5/31/2015.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	UniqueID	GSM2100	GSM2100	Missing	P-Value	Rank	Variance	Num 1.5-Fold	Filter					
6		26	63											
7		-1.98736	-0.01342											
16		-0.63227	-2.0											
18		-1.84377	0.85											
32		0.070188	-2.2											
37		-6.0917	-3											
50		1.601787	-0.3											
59		-3.32617	-1.8											
90		1.890153	-2.2											
114		-4.0917	-1.3											
124		-1.26388	1.16175											
142		1.109934	-0.81237	0										
167		2.176023	-3.23133	0										
172		-1.02131	-3.94753	0										
184		-0.56188	-2.14018	0										
186		-0.46526	-2.77761	0										
191		-0.55565	-2.36257	0										
208		0.066152	1.655564	0										
223		-0.69253	-3.2003	0										
236		-1.72713	-0.04064	0										
269		1.932886	3.671377	0										
288		-1.07147	0.335225	0										

Here I pick up the first one 'UGCluster' for the lookup key. BRB-ArrayTools will start to download required R packages now.

Project - Microsoft Excel

Home Insert Page Layout Formulas Data Review View Add-Ins

ArrayTools  
CGHTools  
Menu Commands

B2

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

A B C D E F G H I J K L M N O P

BRB-ArrayTools

There are 2 arrays in this project. They are shown in the Filtered log ratio/intensity worksheet automatically.

OK

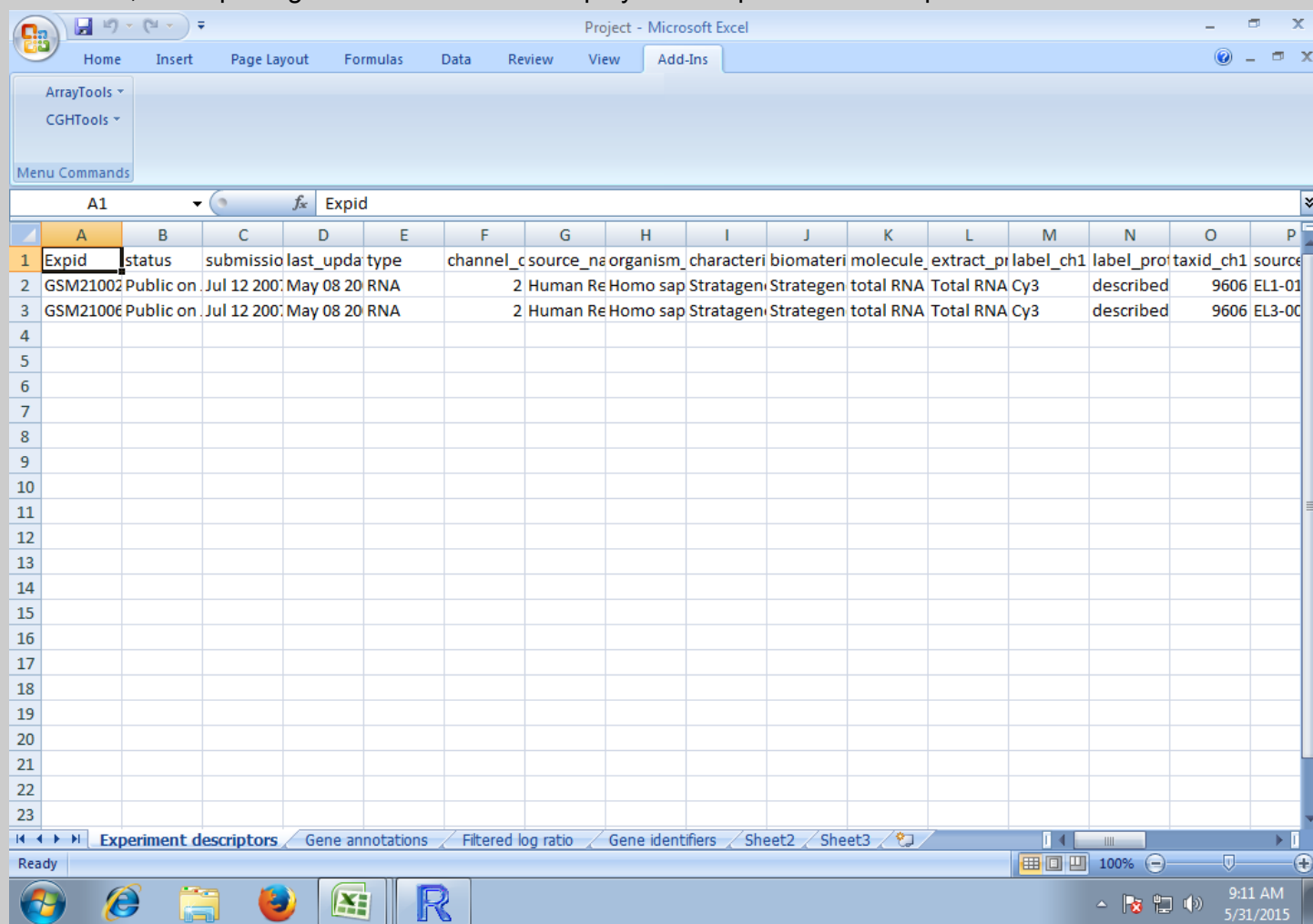
Gene annotations Filtered log ratio Gene identifiers Sheet2 Sheet3

Ready 3192 of 22576 records found

100%

9:10 AM 5/31/2015

At the end, the importing will be finished and displays the 'Experiment descriptors' worksheet.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
	Expid	status	submissio	last_upda	type	channel_c	source_n	organism	characteri	biomateri	molecule	extract_pr	label_ch1	label_pro	taxid_ch1	source
2	GSM21002	Public on	Jul 12 200	May 08 20	RNA	2	Human Re	Homo sap	Stratagen	Strategen	total RNA	Total RNA Cy3		described	9606	EL1-01
3	GSM21006	Public on	Jul 12 200	May 08 20	RNA	2	Human Re	Homo sap	Stratagen	Strategen	total RNA	Total RNA Cy3		described	9606	EL3-00
4																
5																
6																
7																
8																
9																
10																
11																
12																
13																
14																
15																
16																
17																
18																
19																
20																
21																
22																
23																

We can now go around different worksheet to see if everything looks right.

## 'Gene annotations' worksheet.

Project - Microsoft Excel

Home Insert Page Layout Formulas Data Review View Add-Ins

ArrayTools  
CGHTools  
Menu Commands

B2 fx

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
	UniqueID (Double-click)	Name	Accessi	UGClus	Symbol	Entrezl	Chrom	Cytoba	GO	Ensem	Filter					
7	6										TRUE					
17	16										TRUE					
19	18	transmem	NM_00136	Hs.370410	TMCC3	57458	12	12q22	Cellular C	ENSG0000	TRUE					
33	32										TRUE					
38	37										TRUE					
51	50	MYC assoc	NM_00127	Hs.285354	MAX	4149	14	14q23	#####	ENSG0000	TRUE					
60	59	monoacyl	NM_02505	Hs.288568	MOGAT2	80168	11	11q13.5	#####	ENSG0000	TRUE					
91	90	major hist	NM_00551	Hs.77961	HLA-B	3106	6	6p21.3	#####	ENSG0000	TRUE					
115	114	EF-hand d	NM_02518	Hs.521953	EFHC2	80258	X	Xp11.3	Molecular	ENSG0000	TRUE					
125	124	glutamate	NM_00084	Hs.429018	GRM4	2914	6	6p21.3	#####	ENSG0000	TRUE					
143	142										TRUE					
168	167	protein p	NM_00116	Hs.444403	PPP1R12B	4660	1	1q32.1	#####	ENSG0000	TRUE					
173	172	CD300 mo	NM_00116	Hs.147313	CD300LG	146894	17	17q21.31	#####	ENSG0000	TRUE					
185	184										TRUE					
187	186	HECT and	NM_00392	Hs.210385	HERC1	8925	15	15q22	#####	ENSG0000	TRUE					
192	191	zinc finger	NM_00344	Hs.89897	ZNF157	7712	X	Xp11.2	Biological	ENSG0000	TRUE					
209	208	hexokinas	NM_00018	Hs.370365	HK1	3098	10	10q22	#####	ENSG0000	TRUE					
224	223	deiodinas	NM_00136	Hs.49322	DIO3	1735	14	14q32	#####	ENSG0000	TRUE					
237	236	adenylate	NM_02054	Hs.481545	ADCY2	108	5	5p15.3	#####	ENSG0000	TRUE					
270	269	neuropen	NM_00090	Hs.519057	NPY1R	4886	4	4p31.3-p3	#####	ENSG0000	TRUE					

Ready Filter Mode

Experiment descriptors Gene annotations Filtered log ratio Gene identifiers Sheet2 Sheet3

100%

9:13 AM 5/31/2015

## 'Gene identifiers' worksheet.

Project - Microsoft Excel

Home Insert Page Layout Formulas Data Review View Add-Ins

ArrayTools  
CGHTools  
Menu Commands

A1 UniqueID

	A	B	C	D	E	F	G	H	I	J
	UniqueID	Name	Accession	UGClust	Symbol	Ensembl	Entrezl	Defined_Genelists	Filter	
7	6								TRUE	
17	16								TRUE	
19	18	transmembrane and coiled-c	NM_001301	Hs.370410	TMCC3	ENSG00000057704	57458		TRUE	
33	32								TRUE	
38	37								TRUE	
51	50	MYC associated factor X	NM_001271	Hs.285354	MAX	ENSG00000125952	4149	MAPKinase Signaling Pathwa	TRUE	
60	59	monoacylglycerol O-acyltran	NM_025098	Hs.288568	MOGAT2	ENSG00000166391	80168	Fat digestion and absorption	TRUE	
91	90	major histocompatibility com	NM_005514	Hs.77961	HLA-B	ENSG00000234745	3106	Allograft rejection, Antigen p	TRUE	
115	114	EF-hand domain (C-terminal)	NM_025184	Hs.521953	EFHC2	ENSG00000183690	80258		TRUE	
125	124	glutamate receptor, metabot	NM_000841	Hs.429018	GRM4	ENSG00000124493	2914	Neuroactive ligand-receptor	TRUE	
143	142								TRUE	
168	167	protein phosphatase 1, regul	NM_001167	Hs.444403	PPP1R12B	ENSG00000077157	4660	#####	TRUE	
173	172	CD300 molecule-like family r	NM_001168	Hs.147313	CD300LG	ENSG00000161649	146894		TRUE	
185	184								TRUE	
187	186	HECT and RLD domain contain	NM_003922	Hs.210385	HERC1	ENSG00000103657	8925	Ubiquitin mediated proteoly	TRUE	
192	191	zinc finger protein 157	NM_003446	Hs.89897	ZNF157	ENSG00000147117	7712		TRUE	
209	208	hexokinase 1	NM_000188	Hs.370365	HK1	ENSG00000156515	3098	#####	TRUE	
224	223	deiodinase, iodothyronine, t	NM_001362	Hs.49322	DIO3	ENSG00000197406	1735		TRUE	
237	236	adenylate cyclase 2 (brain)	NM_020546	Hs.481545	ADCY2	ENSG00000078295	108	#####	TRUE	
270	269	neuropeptide Y receptor Y1	NM_000909	Hs.519057	NPY1R	ENSG00000164128	4886	Neuroactive ligand-receptor	TRUE	
281	280	biotinidase	NM_000060	Hs.517830	BTD	ENSG00000169814	686	Biotin metabolism, Metaboli	TRUE	

Experiment descriptors Gene annotations Filtered log ratio Gene identifiers Sheet2 Sheet3

Ready Filter Mode 100% 9:15 AM 5/31/2015

'Filtered log ratio' worksheet.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	<a href="#">Display the data</a>	GSM2100 26	GSM2100 63	Missing	P-Value	Rank	Variance	Num 1.5-Fold	Filter					
1	UniqueID													
7	6	-1.98736	-0.01342	0				2	TRUE					
17	16	-0.63227	-2.06839	0				2	TRUE					
19	18	-1.84377	0.854291	0				2	TRUE					
33	32	0.070188	-2.23133	0				2	TRUE					
38	37	-6.0917	-3.2003	0				2	TRUE					
51	50	1.601787	-0.37551	0				2	TRUE					
60	59	-3.32617	-1.81237	0				2	TRUE					
91	90	1.890153	-2.23133	0				2	TRUE					
115	114	-4.0917	-1.38857	0				2	TRUE					
125	124	-1.26388	1.18175	0				2	TRUE					
143	142	1.109934	-0.81237	0				2	TRUE					
168	167	2.176023	-3.23133	0				2	TRUE					
173	172	-1.02131	-3.94753	0				2	TRUE					
185	184	-0.56188	-2.14018	0				2	TRUE					
187	186	-0.46526	-2.77761	0				2	TRUE					
192	191	-0.55565	-2.36257	0				2	TRUE					
209	208	0.066152	1.655564	0				2	TRUE					
224	223	-0.69253	-3.2003	0				2	TRUE					
237	236	-1.72713	-0.04064	0				2	TRUE					
270	269	1.932886	3.671377	0				2	TRUE					

That's it!