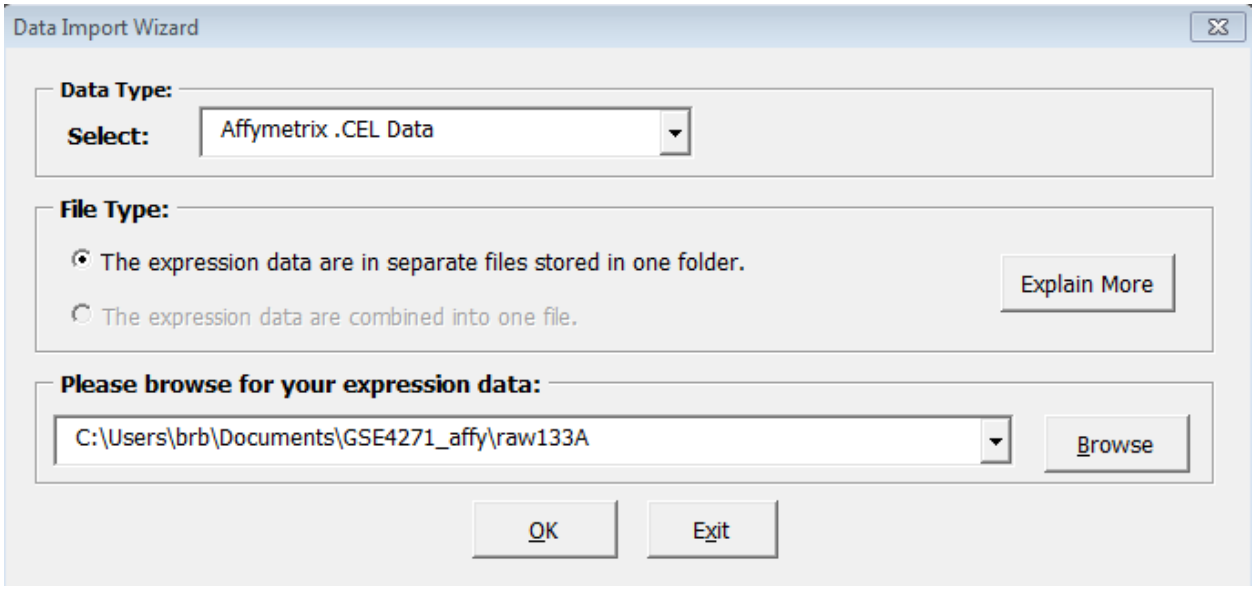
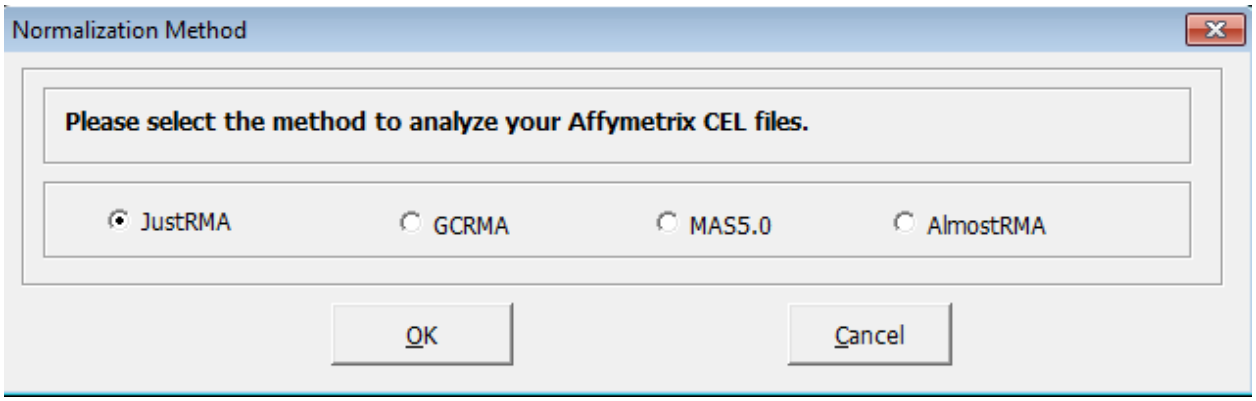
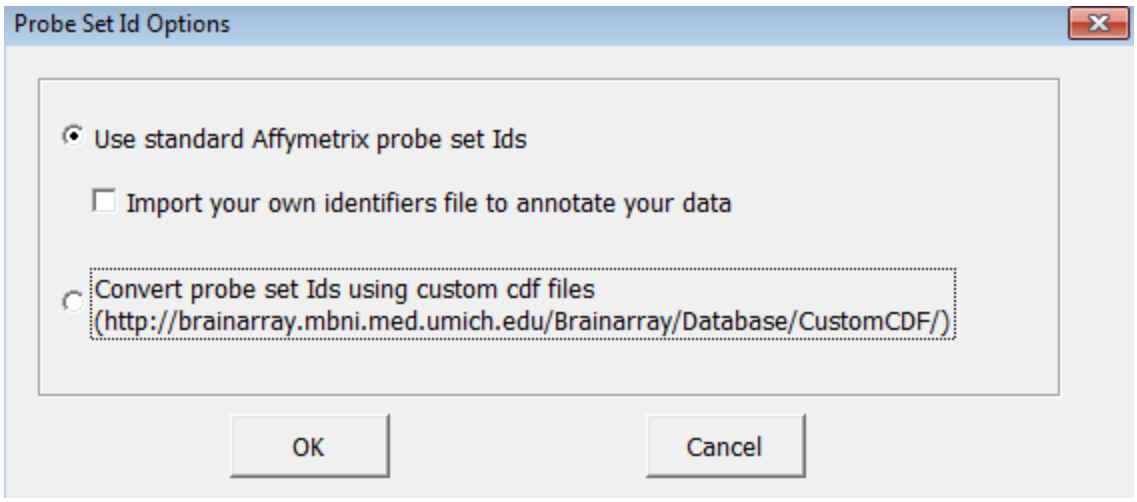
Affy Cel file importer



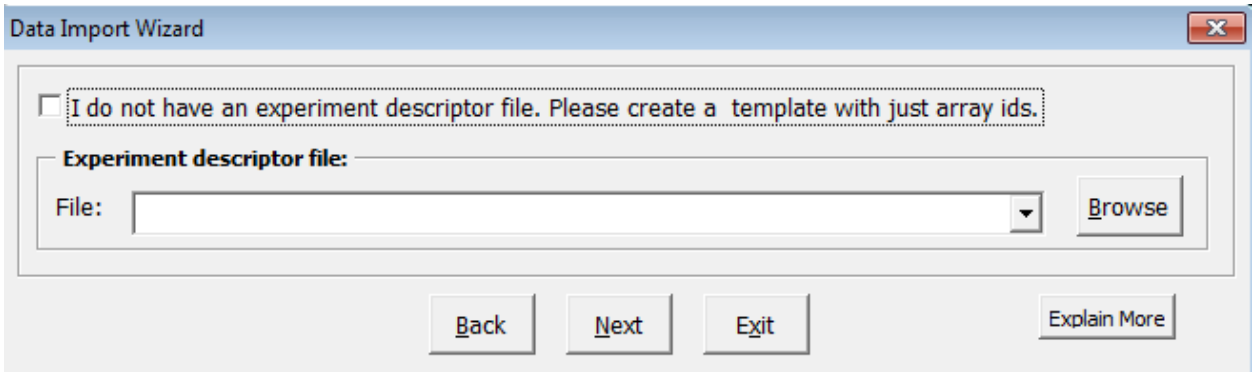


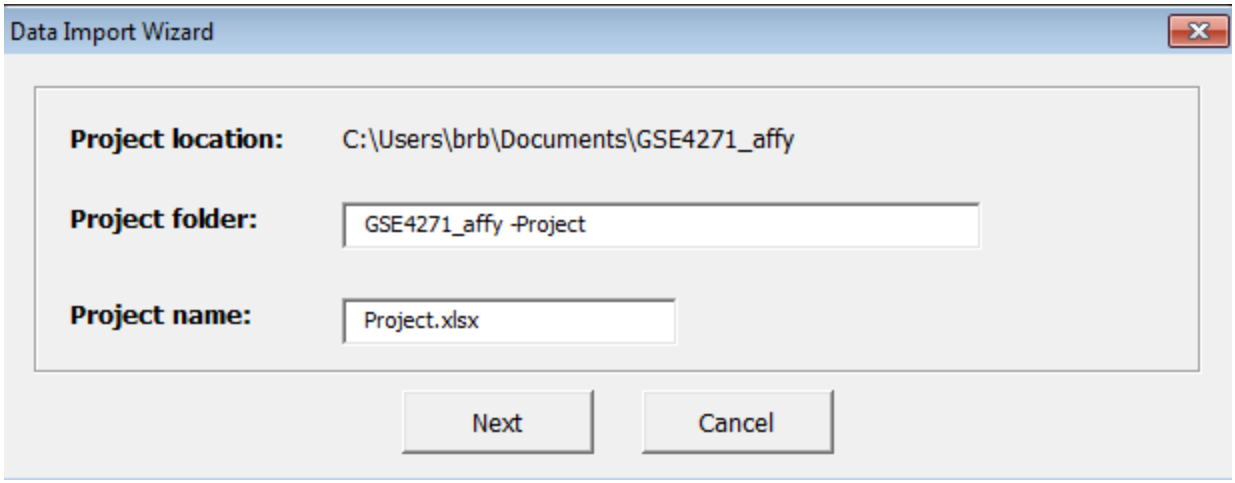
Case 1: Use standard affymetrix probe set id

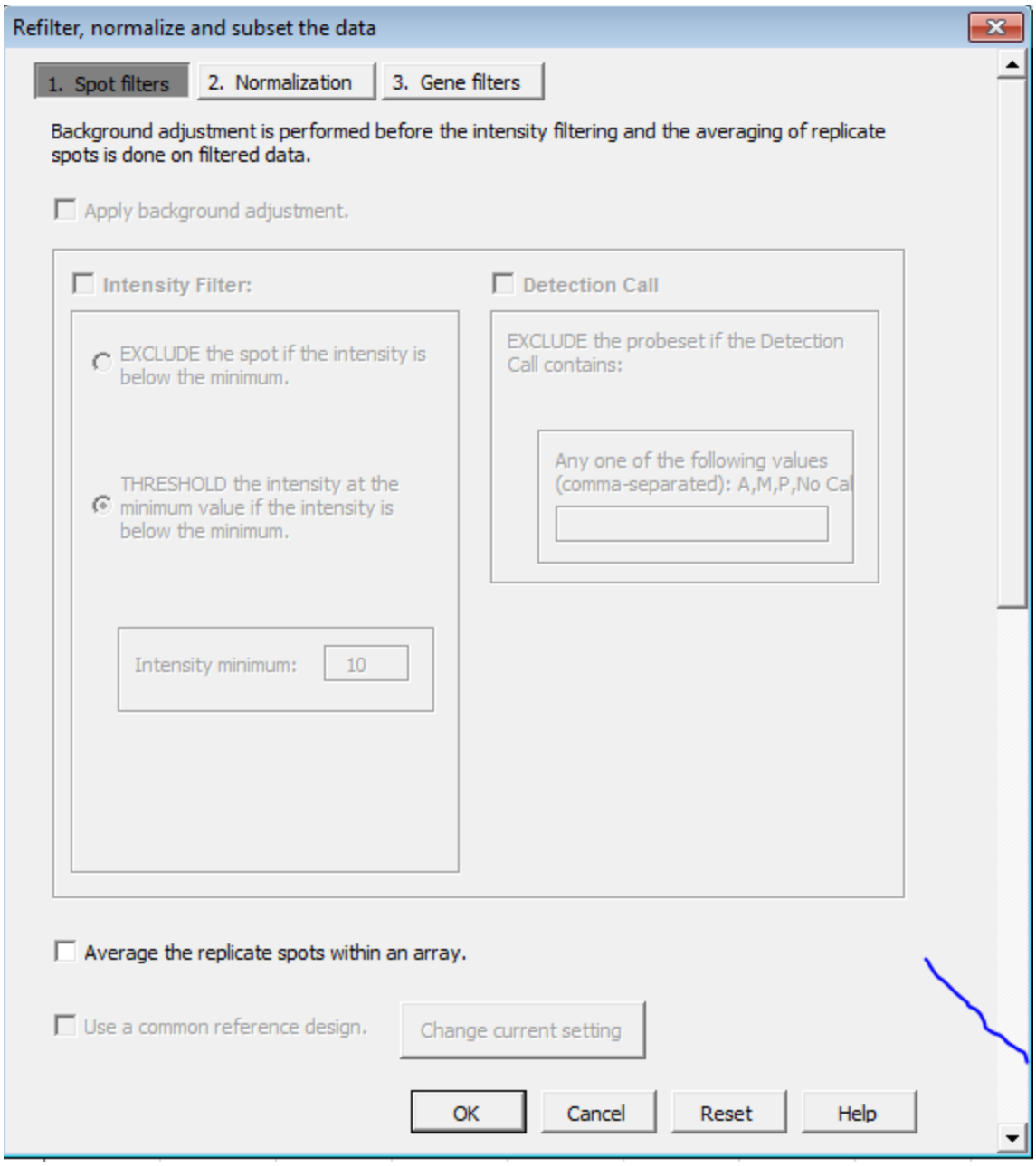


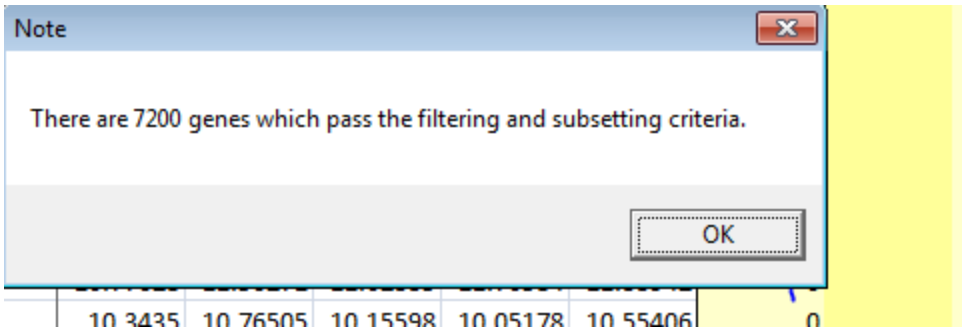
It will install hgu133acdf (only one) package.

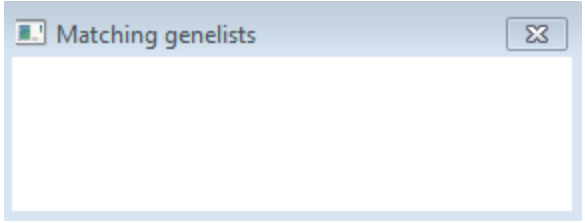
For gcrma option, it will download gcrma package.

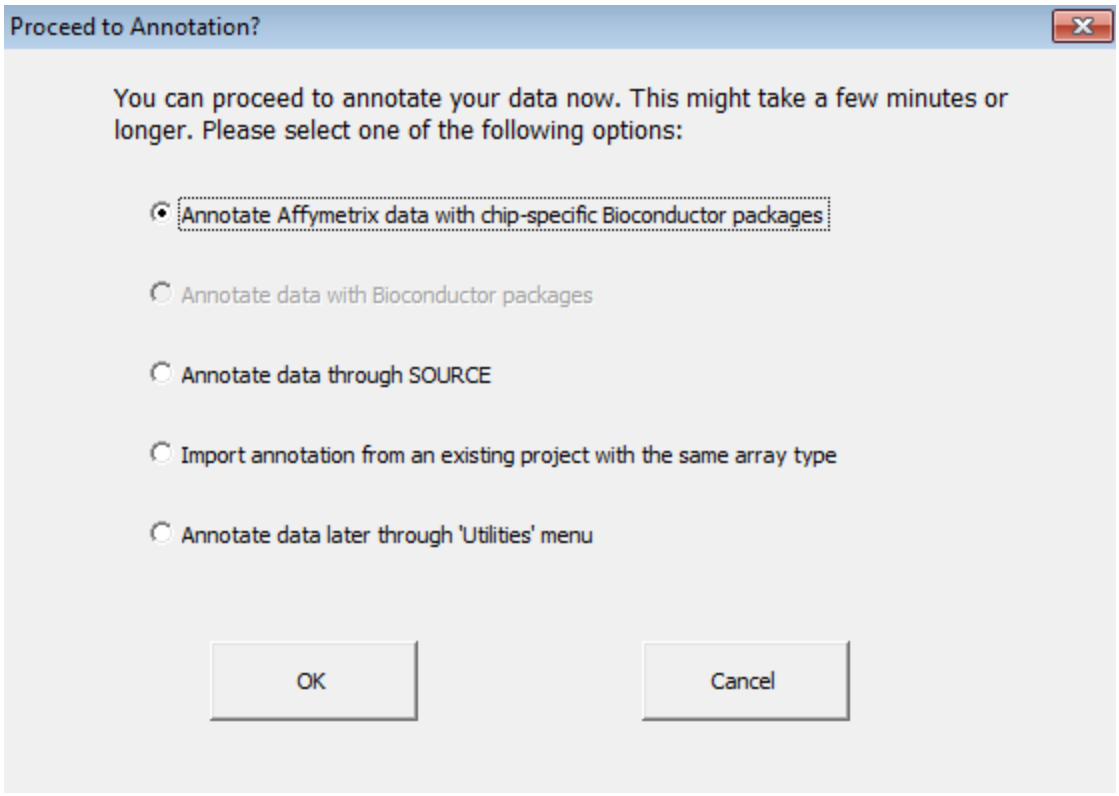




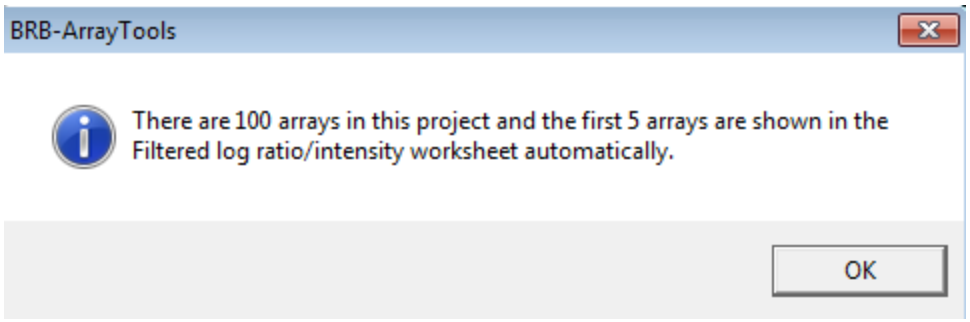

Refilter, normalize and subset the data
1. Spot filters
2. Normalization
3. Gene filters
Background adjustment is performed before the intensity filtering and the averaging of replicate spots is done on filtered data.
Apply background adjustment.
Intensity Filter:
Detection Call
EXCLUDE the spot if the intensity is below the minimum.
THRESHOLD the intensity at the minimum value if the intensity is below the minimum.
EXCLUDE the probeset if the Detection Call contains:
Any one of the following values (comma-separated): A,M,P,No Cal
Intensity minimum:
10
Average the replicate spots within an array.
Use a common reference design.
Change current setting
OK
Cancel
Reset
Help

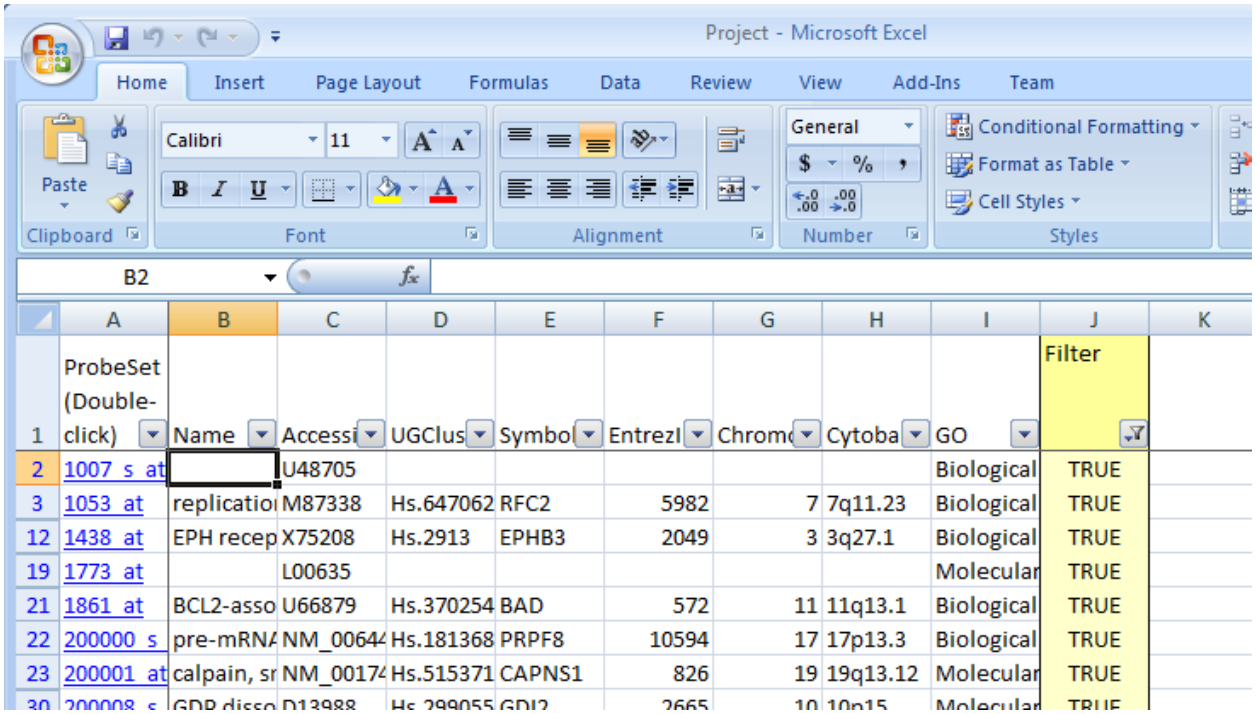

Note
There are 7200 genes which pass the filtering and subsetting criteria.
OK
10.3435
10.76505
10.15598
10.05178
10.55406
0


Matching genelists



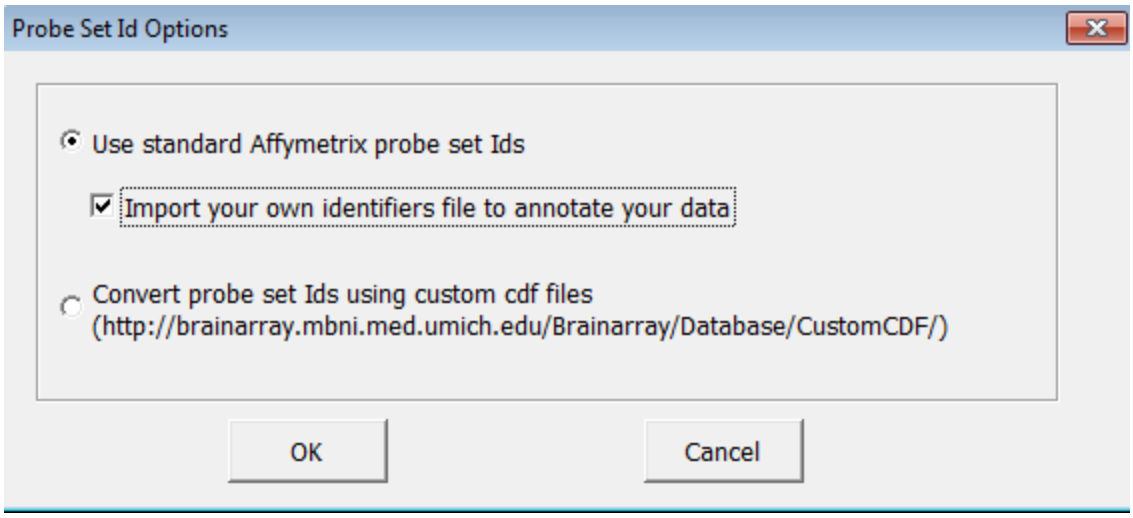
It will install hgu133a.db package.

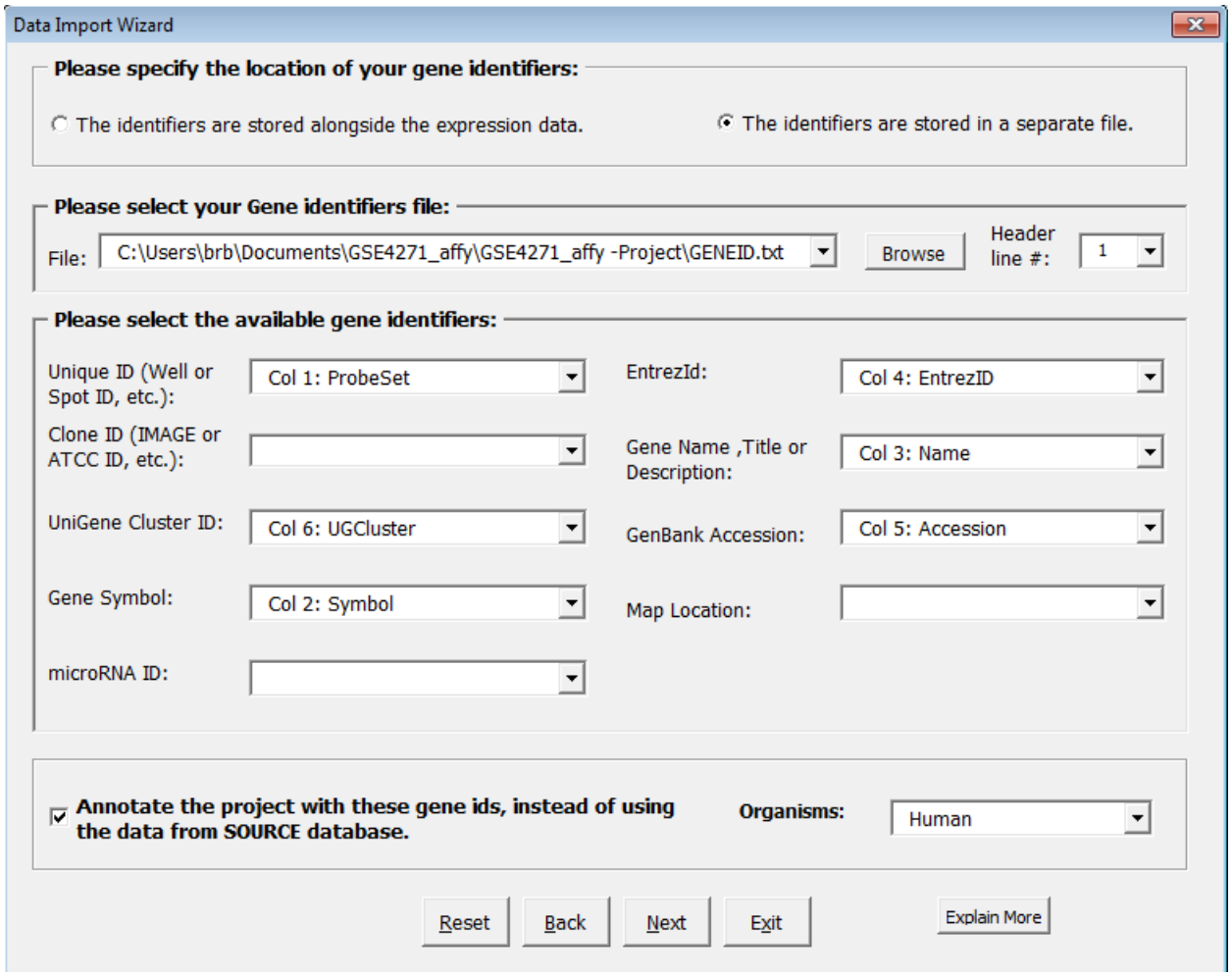


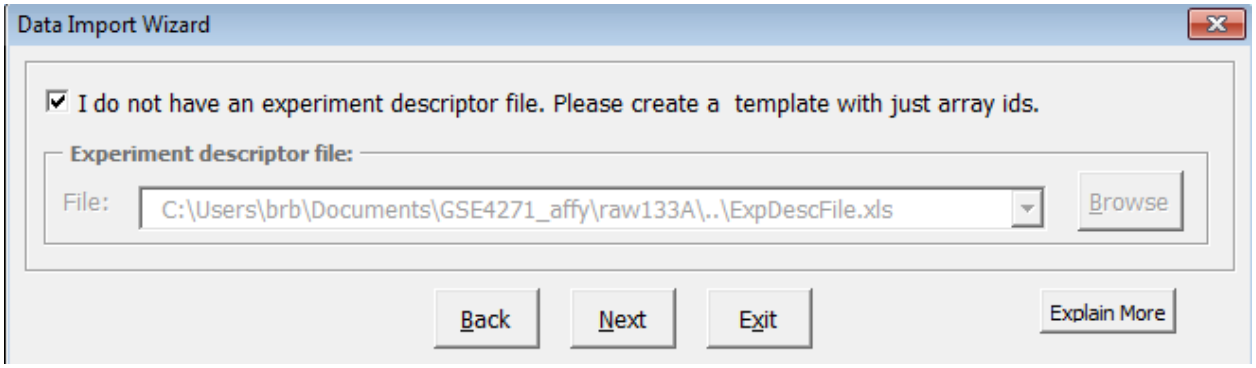


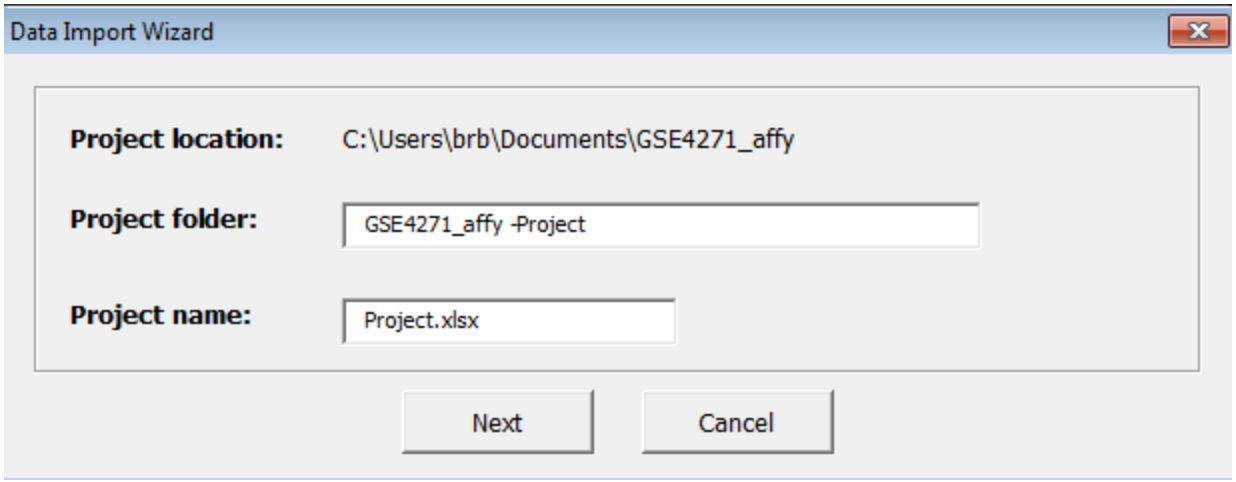
	A	B	C	D	E	F	G	H	I	J	K
1	ProbeSet (Double-click)	Name	Accessi	UGClus	Symbol	Entrezl	Chrom	Cytoba	GO	Filter	
2	1007_s_at		U48705						Biological	TRUE	
3	1053_at	replicatio	M87338	Hs.647062	RFC2	5982	7	7q11.23	Biological	TRUE	
12	1438_at	EPH recep	X75208	Hs.2913	EPHB3	2049	3	3q27.1	Biological	TRUE	
19	1773_at		L00635						Molecular	TRUE	
21	1861_at	BCL2-asso	U66879	Hs.370254	BAD	572	11	11q13.1	Biological	TRUE	
22	200000_s	pre-mRNA	NM_0064	Hs.181368	PRPF8	10594	17	17p13.3	Biological	TRUE	
23	200001_at	calpain, sr	NM_0017	Hs.515371	CAPNS1	826	19	19q13.12	Molecular	TRUE	
30	200008_s	GDP disso	D13988	Hs.299055	GDI2	2665	10	10p15	Molecular	TRUE	

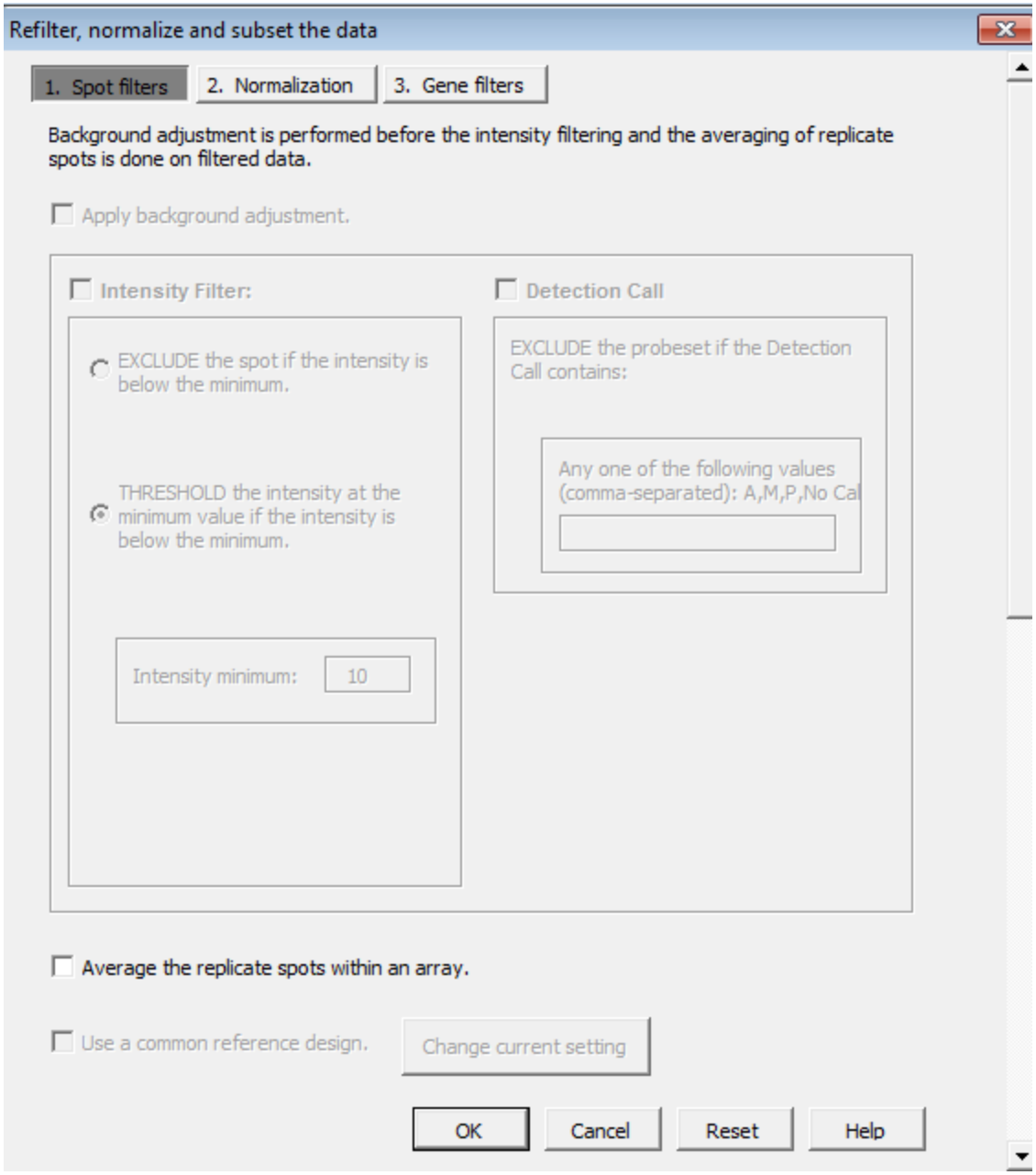
Case 2: Use Affymetrix and import own identifier file

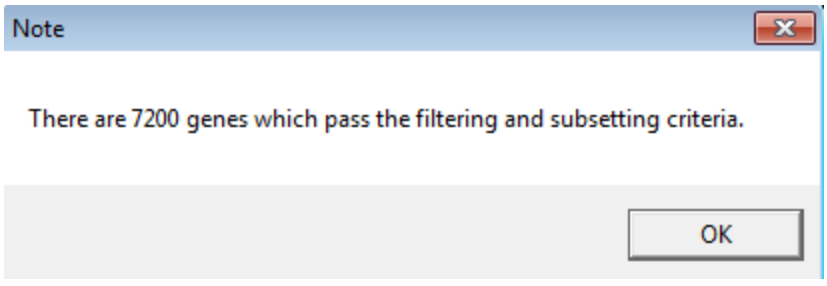


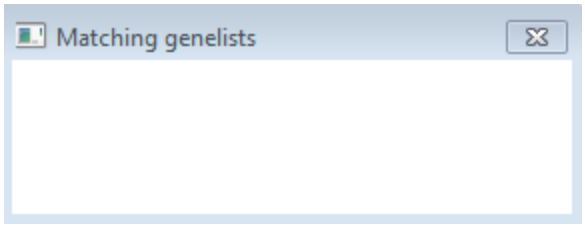

Data Import Wizard
Please specify the location of your gene identifiers:
The identifiers are stored alongside the expression data.
The identifiers are stored in a separate file.
Please select your Gene identifiers file:
File:
C:\Users\brb\Documents\GSE4271_affy\GSE4271_affy -Project\GENEID.txt
Browse
Header line #:
1
Please select the available gene identifiers:
Unique ID (Well or Spot ID, etc.):
Col 1: ProbeSet
EntrezId:
Col 4: EntrezID
Clone ID (IMAGE or ATCC ID, etc.):
Gene Name ,Title or Description:
Col 3: Name
UniGene Cluster ID:
Col 6: UGCluster
GenBank Accession:
Col 5: Accession
Gene Symbol:
Col 2: Symbol
Map Location:
microRNA ID:
Annotate the project with these gene ids, instead of using the data from SOURCE database.
Organisms:
Human
Reset
Back
Next
Exit
Explain More

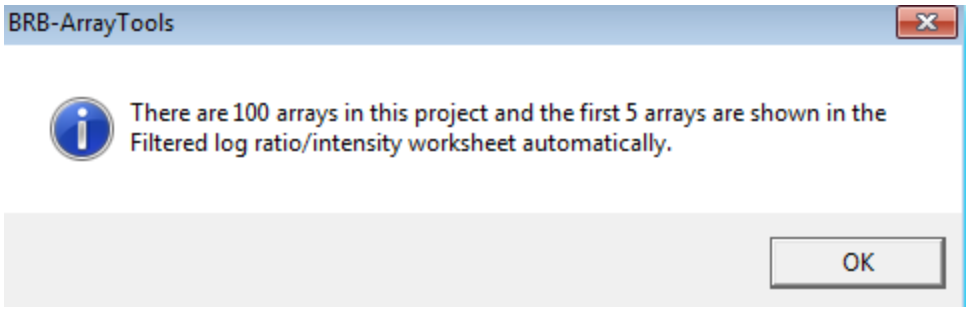

Data Import Wizard
I do not have an experiment descriptor file. Please create a template with just array ids.
Experiment descriptor file:
File:
C:\Users\brb\Documents\GSE4271_affy\raw133A\..\ExpDescFile.xls
Browse
Back
Next
Exit
Explain More

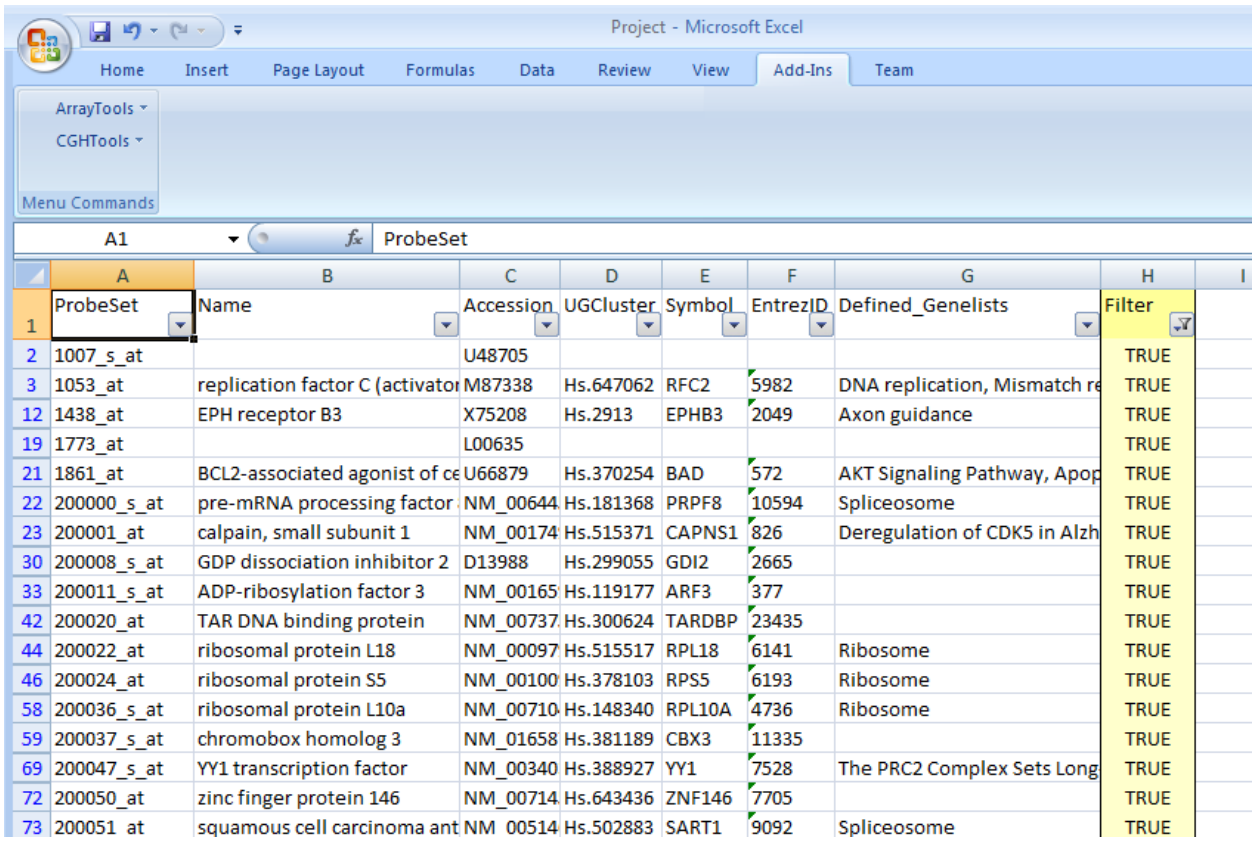

Data Import Wizard
Project location:
C:\Users\brb\Documents\GSE4271_affy
Project folder:
GSE4271_affy -Project
Project name:
Project.xlsx
Next
Cancel


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:
Detection Call
EXCLUDE the spot if the intensity is below the minimum.
THRESHOLD the intensity at the minimum value if the intensity is below the minimum.
Intensity minimum:
10
EXCLUDE the probeset if the Detection Call contains:
Any one of the following values (comma-separated): A,M,P,No Cal
Average the replicate spots within an array.
Use a common reference design.
Change current setting
OK
Cancel
Reset
Help


Note
There are 7200 genes which pass the filtering and subsetting criteria.
OK

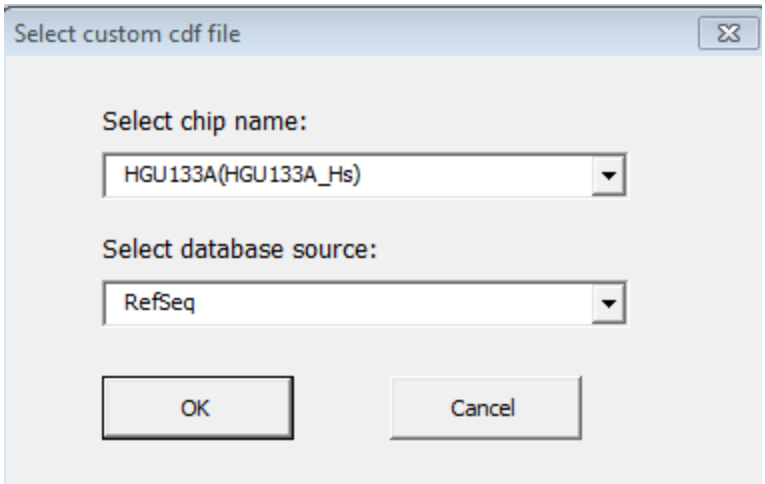

Matching genelists

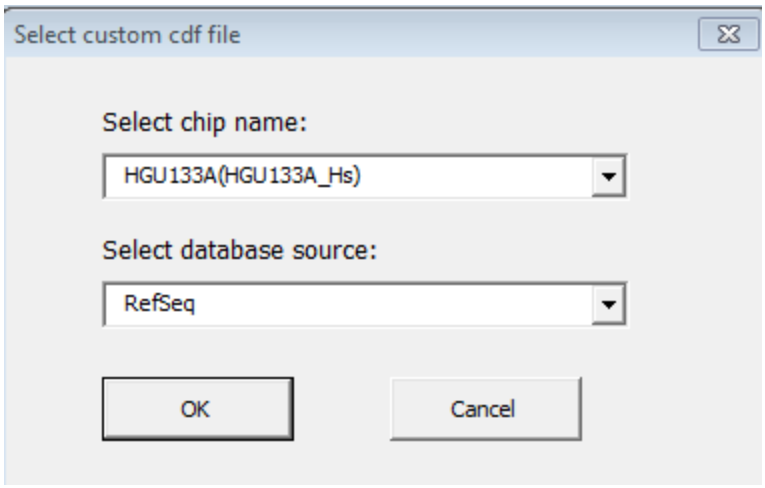




	A	B	C	D	E	F	G	H	I
1	ProbeSet	Name	Accession	UGCluster	Symbol	EntrezID	Defined_Genelists	Filter	
2	1007_s_at		U48705					TRUE	
3	1053_at	replication factor C (activator	M87338	Hs.647062	RFC2	5982	DNA replication, Mismatch re	TRUE	
12	1438_at	EPH receptor B3	X75208	Hs.2913	EPHB3	2049	Axon guidance	TRUE	
19	1773_at		L00635					TRUE	
21	1861_at	BCL2-associated agonist of ce	U66879	Hs.370254	BAD	572	AKT Signaling Pathway, Apop	TRUE	
22	200000_s_at	pre-mRNA processing factor	NM_00644	Hs.181368	PRPF8	10594	Spliceosome	TRUE	
23	200001_at	calpain, small subunit 1	NM_00174	Hs.515371	CAPNS1	826	Deregulation of CDK5 in Alzh	TRUE	
30	200008_s_at	GDP dissociation inhibitor 2	D13988	Hs.299055	GDI2	2665		TRUE	
33	200011_s_at	ADP-ribosylation factor 3	NM_00165	Hs.119177	ARF3	377		TRUE	
42	200020_at	TAR DNA binding protein	NM_00737	Hs.300624	TARDBP	23435		TRUE	
44	200022_at	ribosomal protein L18	NM_00097	Hs.515517	RPL18	6141	Ribosome	TRUE	
46	200024_at	ribosomal protein S5	NM_00100	Hs.378103	RPS5	6193	Ribosome	TRUE	
58	200036_s_at	ribosomal protein L10a	NM_00710	Hs.148340	RPL10A	4736	Ribosome	TRUE	
59	200037_s_at	chromobox homolog 3	NM_01658	Hs.381189	CBX3	11335		TRUE	
69	200047_s_at	YY1 transcription factor	NM_00340	Hs.388927	YY1	7528	The PRC2 Complex Sets Long	TRUE	
72	200050_at	zinc finger protein 146	NM_00714	Hs.643436	ZNF146	7705		TRUE	
73	200051_at	squamous cell carcinoma ant	NM_00514	Hs.502883	SART1	9092	Spliceosome	TRUE	

Case 3: Use custom cdf

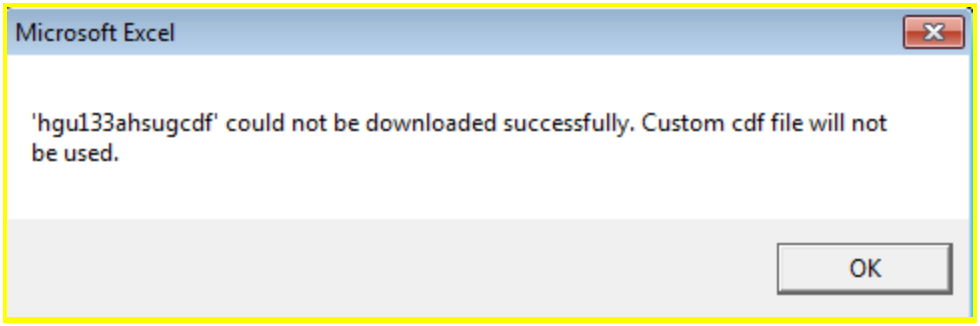


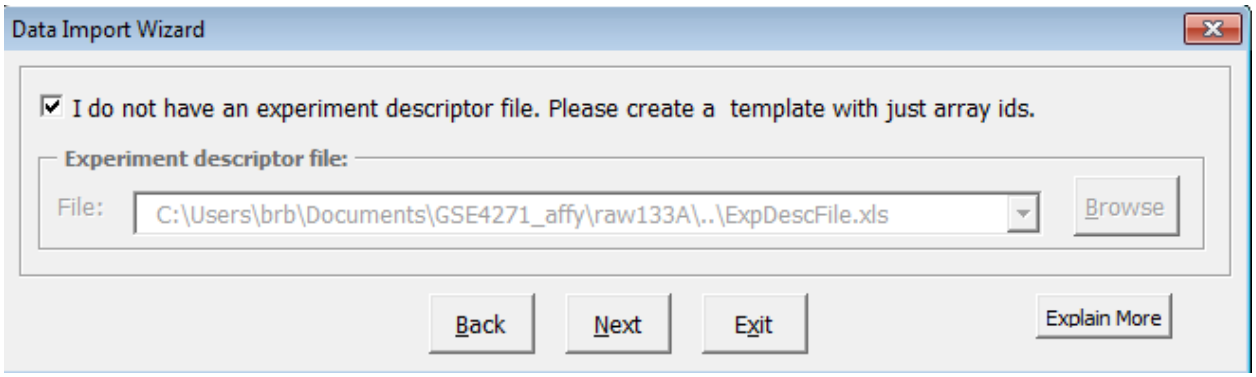


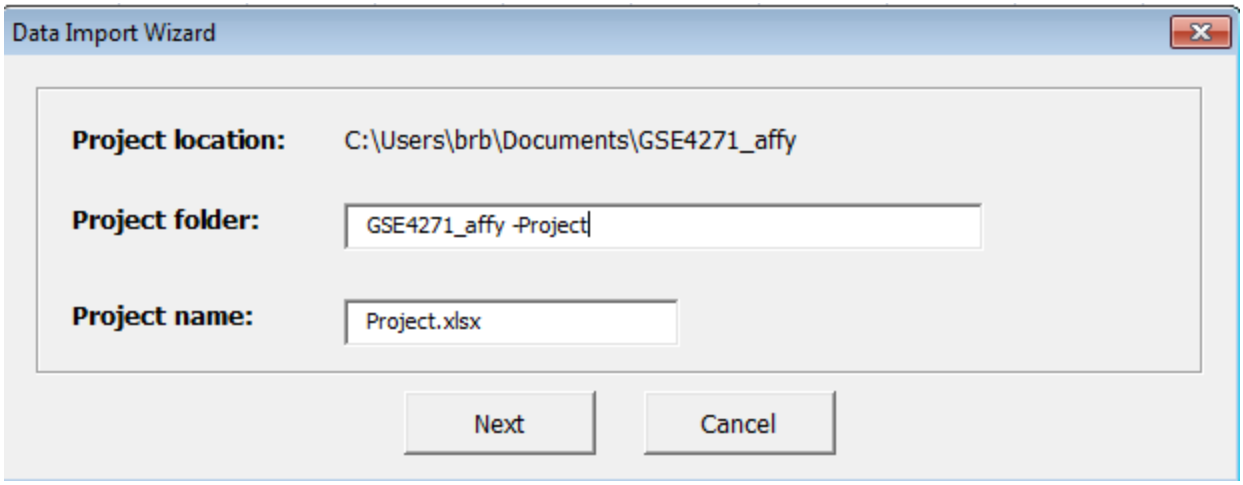
database source: EntrezGene, RefSeq, UniGene

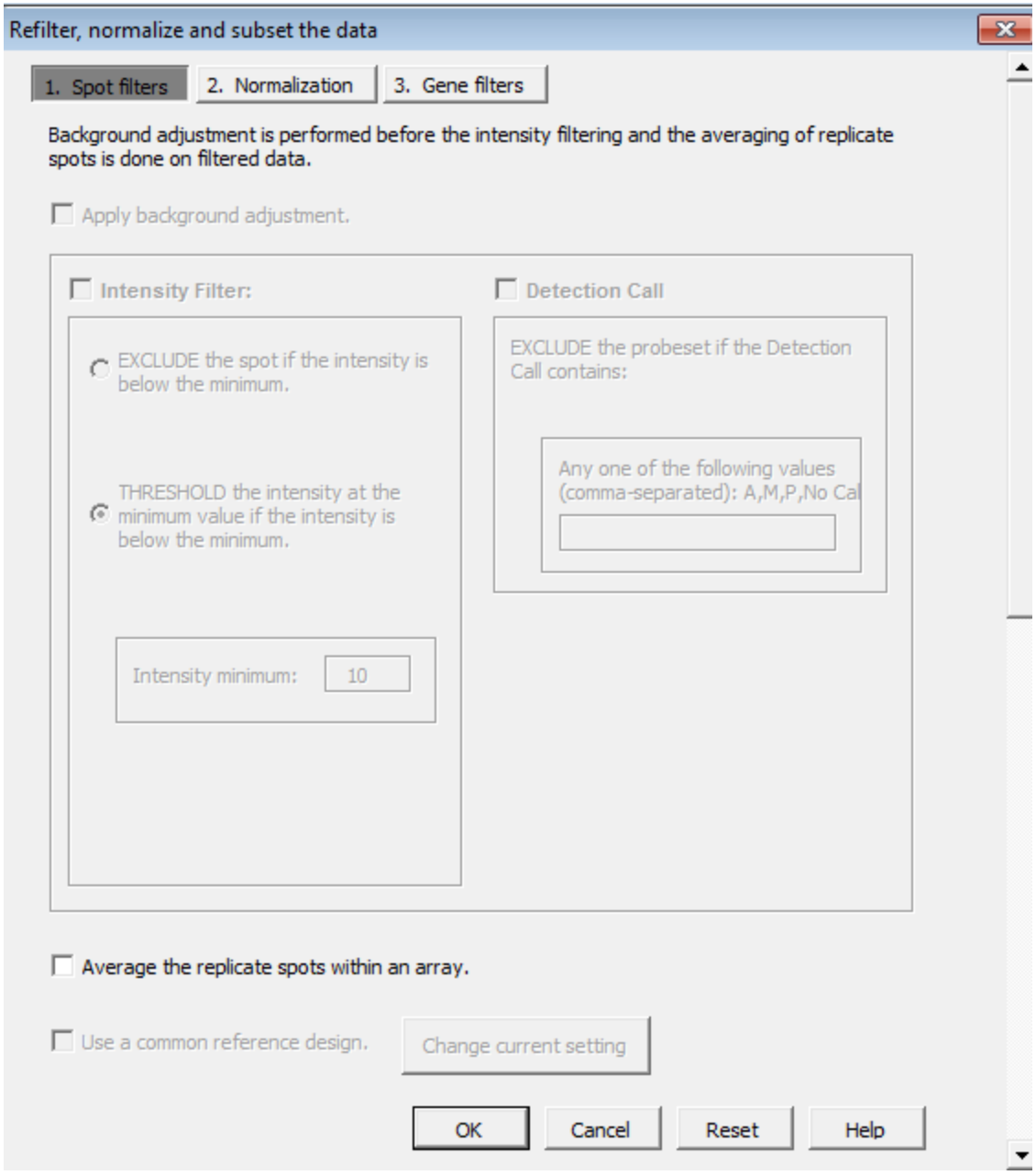
It will download hgu133ahsentrezgprobe, hgu133ahsentrezg.db, hgu133ahsentrezgcdf together if we choose RefSeq.

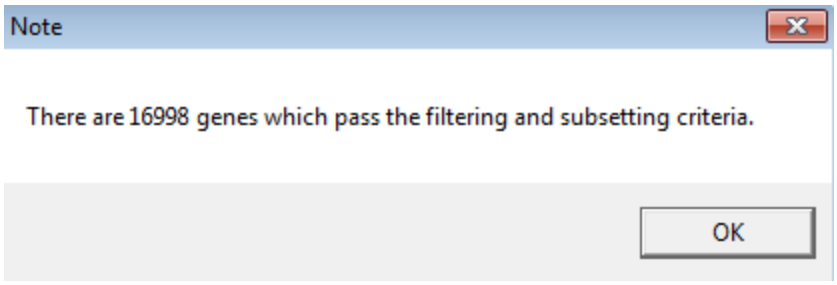
it will download hgu133ahsugcdf if we choose UniGene

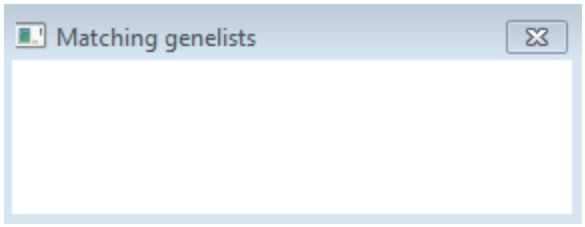


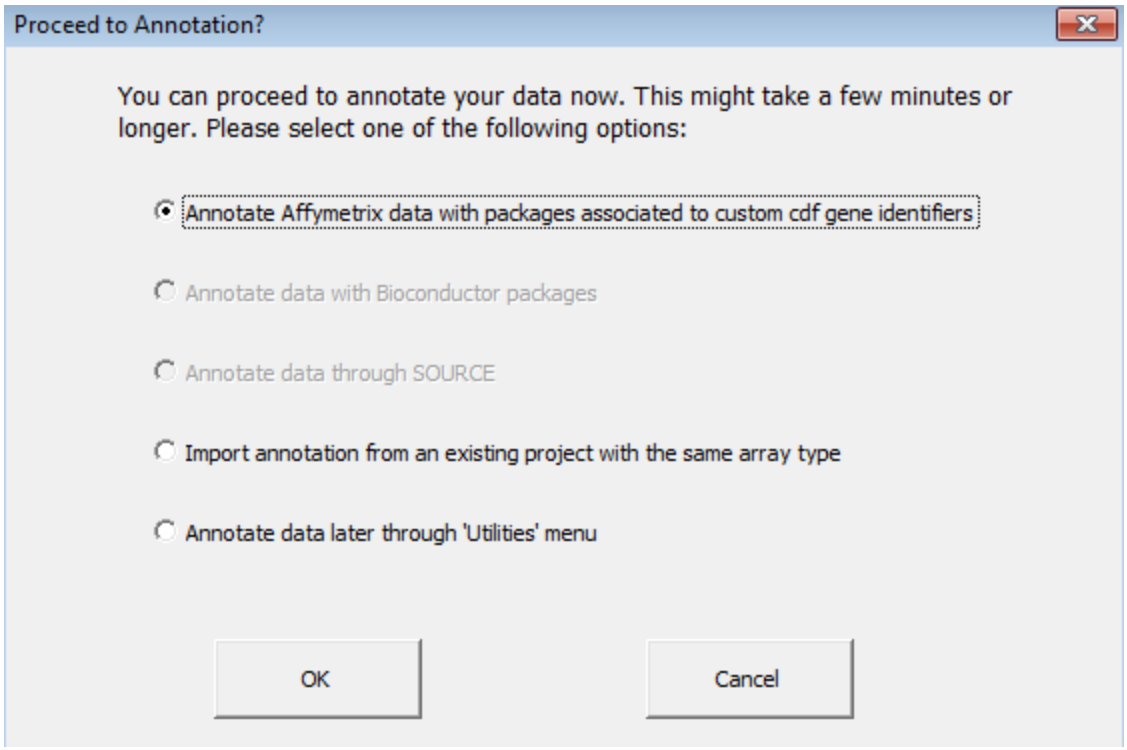


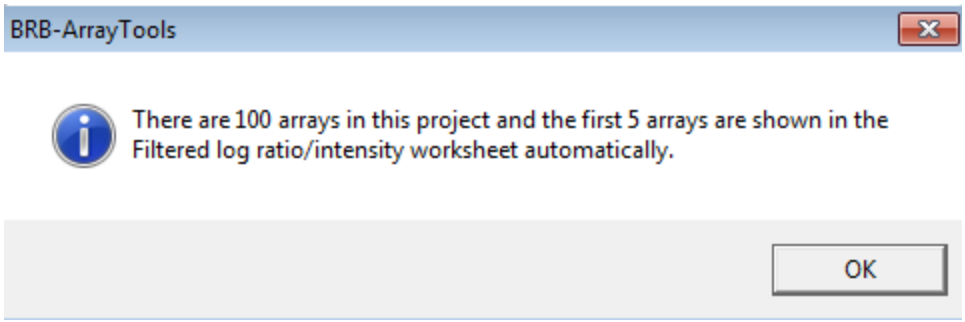

Data Import Wizard
Project location:
C:\Users\brb\Documents\GSE4271_affy
Project folder:
GSE4271_affy -Project
Project name:
Project.xlsx
Next
Cancel

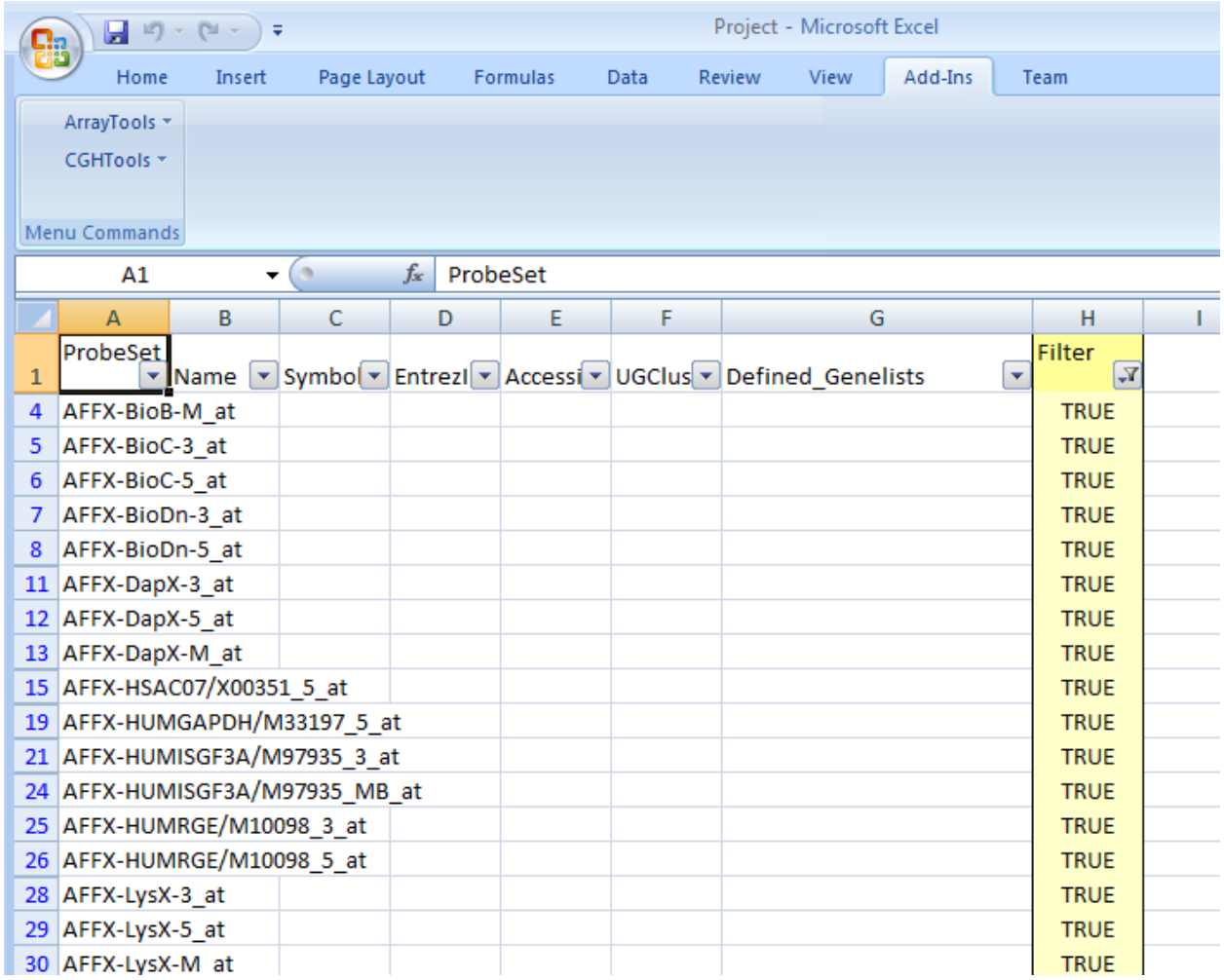

Refilter, normalize and subset the data
1. Spot filters
2. Normalization
3. Gene filters
Background adjustment is performed before the intensity filtering and the averaging of replicate spots is done on filtered data.
Apply background adjustment.
Intensity Filter:
EXCLUDE the spot if the intensity is below the minimum.
THRESHOLD the intensity at the minimum value if the intensity is below the minimum.
Intensity minimum:
10
Detection Call
EXCLUDE the probeset if the Detection Call contains:
Any one of the following values (comma-separated): A,M,P,No Cal
Average the replicate spots within an array.
Use a common reference design.
Change current setting
OK
Cancel
Reset
Help


Note
There are 16998 genes which pass the filtering and subsetting criteria.
OK


Matching genelists


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 packages associated to custom cdf gene identifiers
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


BRB-ArrayTools
There are 100 arrays in this project and the first 5 arrays are shown in the Filtered log ratio/intensity worksheet automatically.
OK



	A	B	C	D	E	F	G	H	I
1	ProbeSet	Name	Symbo	EntrezI	Accessi	UGClus	Defined_Genelists	Filter	
4	AFFX-BioB-M_at							TRUE	
5	AFFX-BioC-3_at							TRUE	
6	AFFX-BioC-5_at							TRUE	
7	AFFX-BioDn-3_at							TRUE	
8	AFFX-BioDn-5_at							TRUE	
11	AFFX-DapX-3_at							TRUE	
12	AFFX-DapX-5_at							TRUE	
13	AFFX-DapX-M_at							TRUE	
15	AFFX-HSAC07/X00351_5_at							TRUE	
19	AFFX-HUMGAPDH/M33197_5_at							TRUE	
21	AFFX-HUMISGF3A/M97935_3_at							TRUE	
24	AFFX-HUMISGF3A/M97935_MB_at							TRUE	
25	AFFX-HUMRGE/M10098_3_at							TRUE	
26	AFFX-HUMRGE/M10098_5_at							TRUE	
28	AFFX-LysX-3_at							TRUE	
29	AFFX-LysX-5_at							TRUE	
30	AFFX-LysX-M_at							TRUE	