

caArray 002 - Uploading MicroArray Gene Expression Data into caArray

Question: How do I upload MicroArray Gene Expression Data into caArray?

Topic: caArray Usage

Release: caArray 2.0 and above

Date entered: 01/06/2009

I want to load data related to a microarray gene expression experiment into caArray. Where should each piece of data go in caArray?

Answer

A microarray gene expression experiment is typically associated with the following types of files: array design files, sample annotation files, and raw and/or processed data files. In addition, a user might also want to associate with a particular experiment, files that describe the experiment, files that describe the samples, and other miscellaneous files.

caArray allows a user to load all types of files into the system. Furthermore, caArray performs data parsing for some of those files if it has the corresponding data parsing modules installed. The [caArray User's Guide](#) has detailed descriptions of the types of files supported by caArray and steps taken to upload, validate, and import those files. See [About File Types](#).

Given that in most cases the user uploads data generated from one particular array platform at a time, this document provides a step-wise description of the data upload/validate/import process in caArray for each supported array platform.

Step 1. (For all array platforms): Check the availability of Array Design File(s)

Before an experiment is created in caArray, you should first check whether the corresponding Array Design File(s) already exist in caArray. To do so, select **Manage Array Designs** under **Curation** on the left sidebar, as shown in the image below.

Manage Array Designs Screen Listing the Design Files

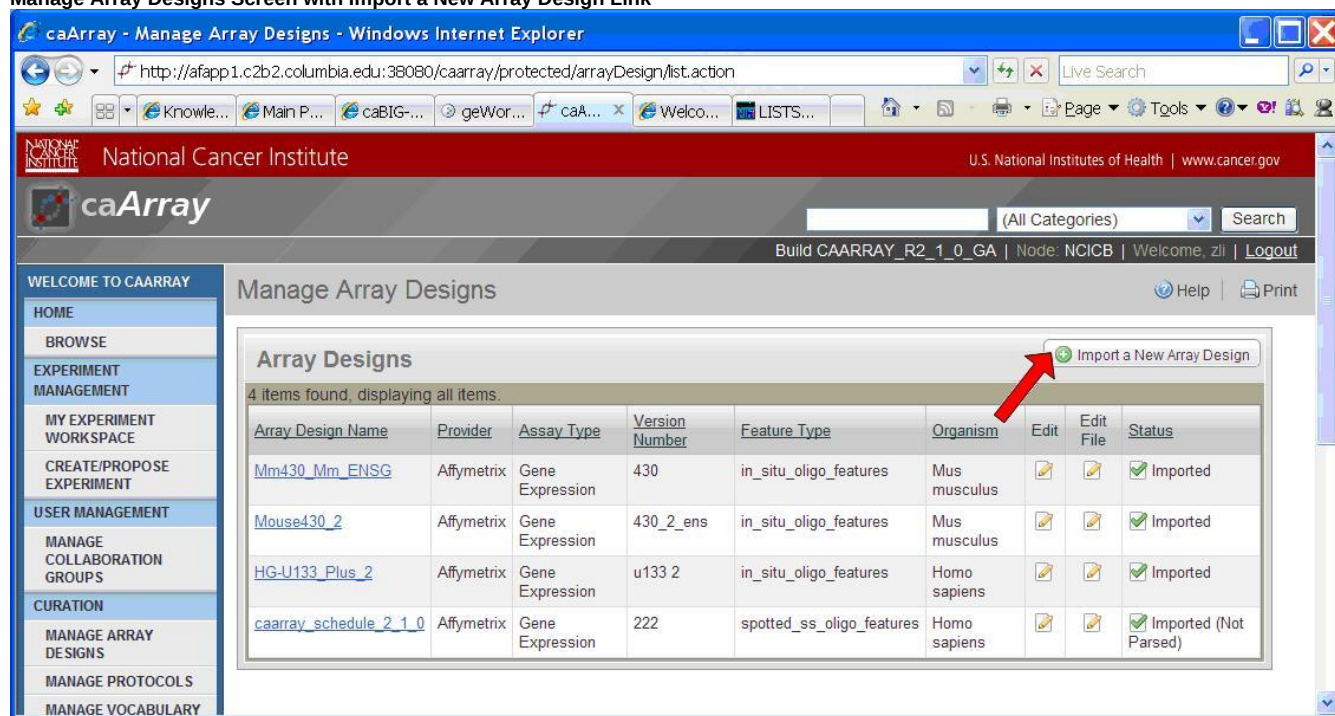
The screenshot shows the 'Manage Array Designs' screen in the caArray application. The left sidebar contains a navigation menu with the following items: WELCOME TO CAARRAY, HOME, BROWSE, EXPERIMENT MANAGEMENT, MY EXPERIMENT WORKSPACE, CREATE/PROPOSE EXPERIMENT, USER MANAGEMENT, MANAGE COLLABORATION GROUPS, CURATION (highlighted), MANAGE ARRAY DESIGNS (highlighted with a red arrow), MANAGE PROTOCOLS, and MANAGE VOCABULARY. The main content area is titled 'Manage Array Designs' and shows a table of 4 array designs. The table has columns: Array Design Name, Provider, Assay Type, Version Number, Feature Type, Organism, Edit, Edit File, and Status. The designs listed are: Mm430_Mm_ENSG, Mouse430_2, HG-U133_Plus_2, and caarray_schedule_2_1_0. All designs are marked as 'Imported'.

Array Design Name	Provider	Assay Type	Version Number	Feature Type	Organism	Edit	Edit File	Status
Mm430_Mm_ENSG	Affymetrix	Gene Expression	430	in_situ_oligo_features	Mus musculus			Imported
Mouse430_2	Affymetrix	Gene Expression	430_2_ens	in_situ_oligo_features	Mus musculus			Imported
HG-U133_Plus_2	Affymetrix	Gene Expression	u133_2	in_situ_oligo_features	Homo sapiens			Imported
caarray_schedule_2_1_0	Affymetrix	Gene Expression	222	spotted_ss_oligo_features	Homo sapiens			Imported (Not Parsed)

Figure 1. Check available array designs in caArray.

If the corresponding Array Design File(s) for the array platform used in the experiment are found, proceed to Step 2.

Manage Array Designs Screen with Import a New Array Design Link



The screenshot shows the 'Manage Array Designs' screen in the caArray application. The browser window title is 'caArray - Manage Array Designs - Windows Internet Explorer'. The URL is 'http://afapp1.c2b2.columbia.edu:38080/caarray/protected/arrayDesign/list.action'. The page header includes the National Cancer Institute logo and the text 'National Cancer Institute' and 'U.S. National Institutes of Health | www.cancer.gov'. The caArray logo is also present. A search bar is located at the top right. The main content area is titled 'Manage Array Designs' and contains a table of array designs. A red arrow points to the 'Import a New Array Design' link in the top right corner of the table area.

Array Designs

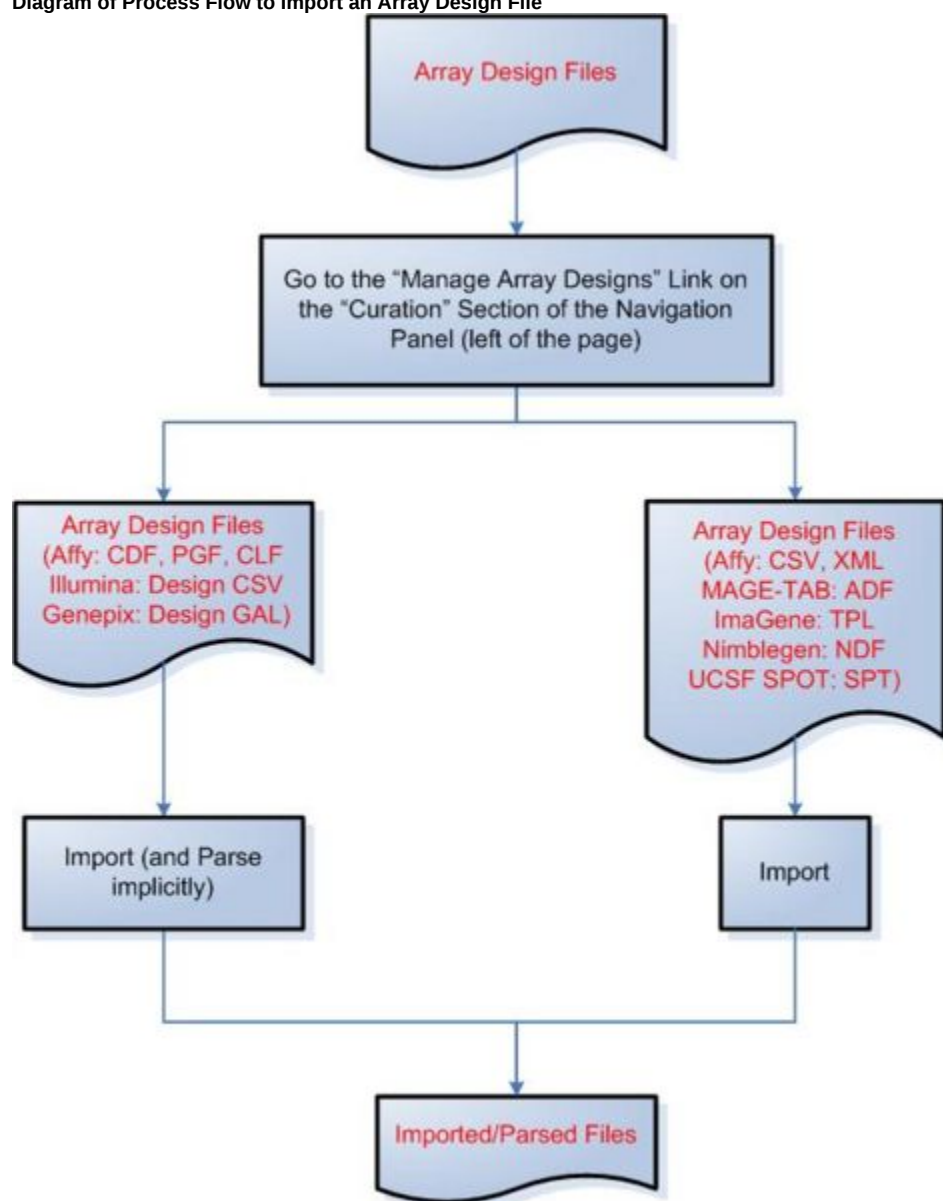
4 items found, displaying all items.

Array Design Name	Provider	Assay Type	Version Number	Feature Type	Organism	Edit	Edit File	Status
Mm430_Mm_ENSG	Affymetrix	Gene Expression	430	in_situ_oligo_features	Mus musculus			Imported
Mouse430_2	Affymetrix	Gene Expression	430_2_ens	in_situ_oligo_features	Mus musculus			Imported
HG-U133_Plus_2	Affymetrix	Gene Expression	u133 2	in_situ_oligo_features	Homo sapiens			Imported
caarray_schedule_2_1_0	Affymetrix	Gene Expression	222	spotted_ss_oligo_features	Homo sapiens			Imported (Not Parsed)

Figure 2. Import a new array design.

If they are not there, you should click **Import a New Array Design** as shown in the preceding image, and follow the process described in the diagram below to import an Array Design File. (Only a registered user can import an Array Design File).

Diagram of Process Flow to Import an Array Design File



As shown in the figure above, to begin importing a file an array design file into caArray, first select **Manage Array Designs** under **Curation** in the left sidebar. Then select **Import a New Array Design** button at the top right of the Manage Array Designs page. The following file formats can be both imported and parsed by caArray: Affymetrix CDF, PGF, and CLF; Illumina Design CSV; and Genepix Design GAL. The following file formats can be imported, but not parsed: Affymetrix CSV, XML MAGE-TAB, and ADF; ImaGene TPL; Nimblegen NDF; and UCSF SPOT SPT.

Note that any array design format not mentioned in the flow diagram is not supported by caArray up to version 2.1.1.

Step 2 (for all array platforms): Create an experiment

The [caArray User's Guide](#) provides detailed information for [creating an experiment](#).

Step 3: Upload, validate, and import files into caArray

In this step, slightly different processes are used to load data into caArray for different array platforms.

- [#Step 3A Import data under the MAGE-TAB format](#)
- [#Step 3B Import experimental data using the Affymetrix platform](#)
- [#Step 3C Import experimental data using the Illumina platform](#)
- [#Step 3D Import experimental data using the Genepix platform](#)
- [#Step 3E Import experimental data using the Agilent platform](#)
- [#Step 3F Import experimental data using the ImaGene platform](#)
- [#Step 3G Import experimental data using the Nimblegen platform](#)

Step 3A Import data under the MAGE-TAB format

Only file types listed in the flow diagram below can be imported.

Diagram of Process Flow to Import a MAGE-TAB File

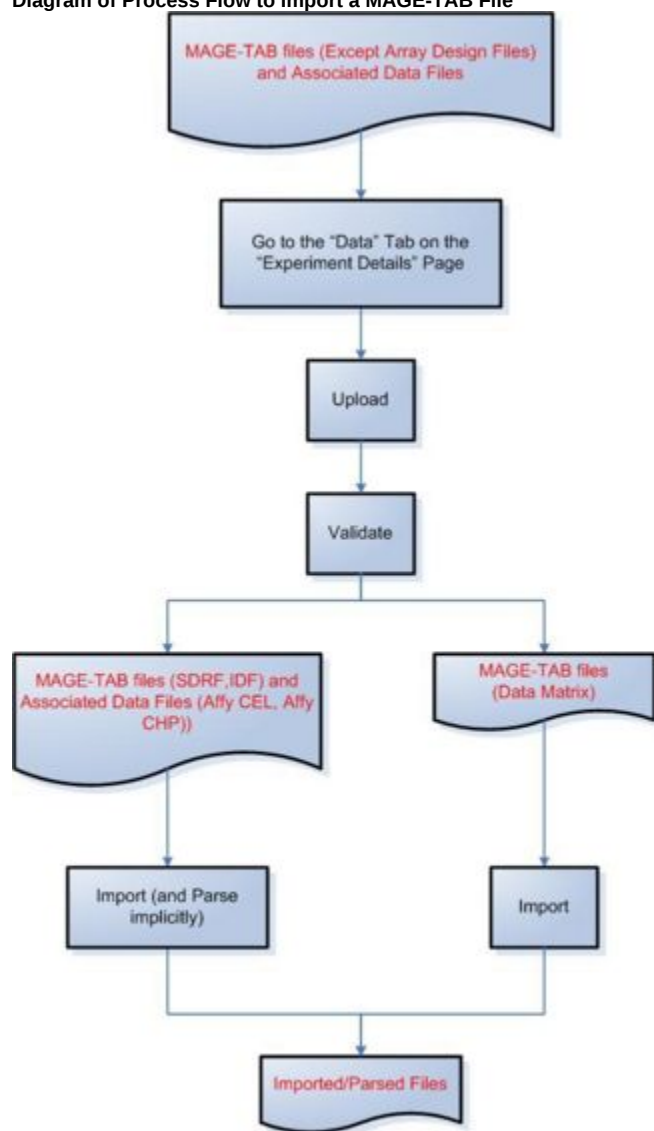


Figure 4. Import data under the MAGE-TAB format

According to the figure above, to import a MAGE-TAB file, first select the **Data** tab on the **Experiment Details** page, then proceed to upload and validate the file you wish to import. The following MAGE-TAB file formats can be both imported and parsed by caArray: SDRF and IDF, along with their associated data files (Affymetrix CEL and CHP). MAGE-TAB data matrix file formats can be imported, but not parsed.

Step 3B Import experimental data using the Affymetrix platform

Because of a known issue concerning importing Affymetrix DAT files, uploaded Affymetrix DAT files should be kept in the "Manage Data" tab under the "Data" tab for a selected experiment for caArray versions up to version 2.1.1. Once the issue is resolved, Affymetrix DAT files will be able to be imported the same way as EXP, RPT, and TXT files.

Diagram of Process Flow to Import an Affymetrix DAT File

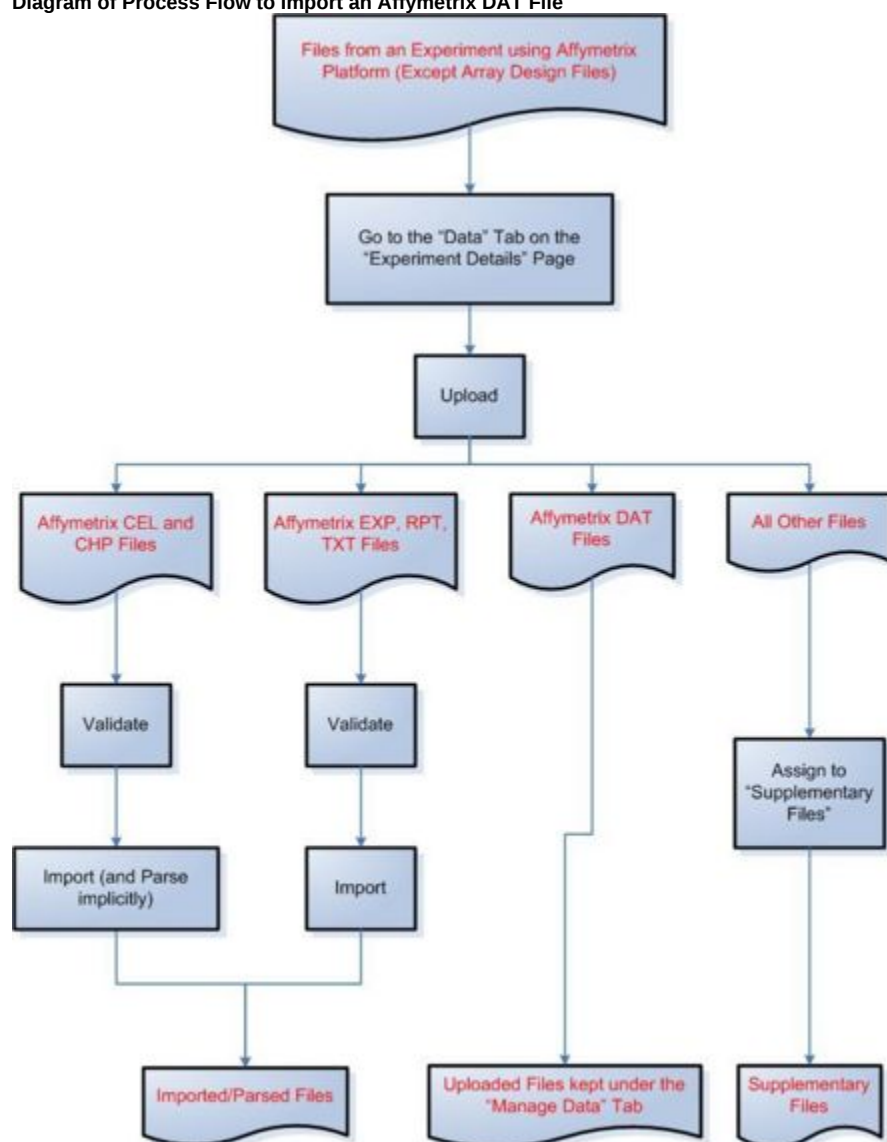


Figure 5. Import experimental data using the Affymetrix platform.

According to the figure above, to import files containing experimental data from the Affymetrix platform, first select the **Data** tab on the **Experiment Details** page, then proceed to upload the file. For a CEL or CHP file, you should validate the file, then import and implicitly parse it. For an EXP, RPT, or TXT file, you should validate the file, then import it. DAT files should not be validated or imported; instead they should be kept under the Manage Data tab. You should assign all other file formats to the 'Supplementary Files' category.

Step 3C Import experimental data using the Illumina platform

Diagram of Process Flow to Import an Illumina Data File

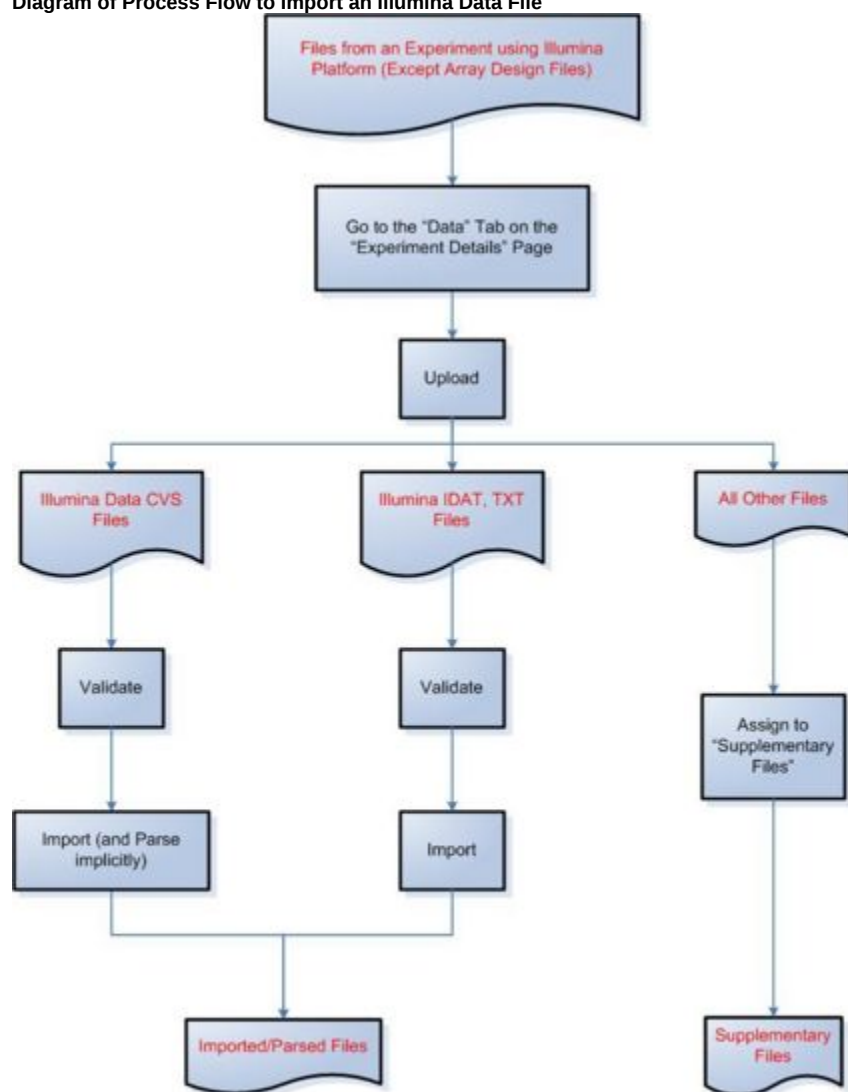


Figure 6. Import experimental data using the Illumina platform.

According to the figure above, to import files containing experimental data from the Illumina platform, first select the **Data** tab on the 'Experiment Details' page, then proceed to upload the file. For a Data CVS file, you should validate the file, then import and implicitly parse it. For an IDAT or TXT file, you should validate the file, then import it. You should assign all other file formats to the **Supplementary Files** category.

Step 3D Import experimental data using the Genepix platform

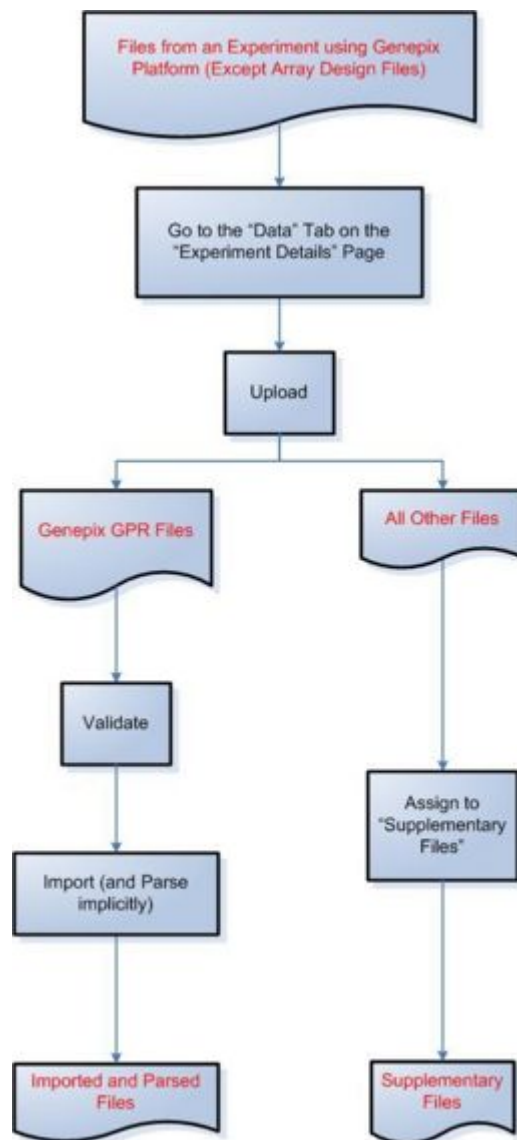


Figure 7. Import experimental data using the Genepix platform.

Diagram of Process Flow to Import a Genepix Data File

According to the figure above, to import files containing experimental data from the Genepix platform, first select the **Data** tab on the **Experiment Details** page, then proceed to upload the file. For a GPR file, you should validate the file, then import and implicitly parse it. You should assign all other file formats to the **Supplementary Files** category.

Step 3E Import experimental data using the Agilent platform

Because of a known issue concerning importing Agilent TSV files, uploaded Agilent TSV files should remain in the "Manage Data" tab under the "Data" tab for a selected experiment for caArray versions up to version 2.1.1. Once the issue is resolved, Agilent TSV files will be able to be imported the same way as TXT files.

Diagram of Process Flow to Import an Agilent Data File

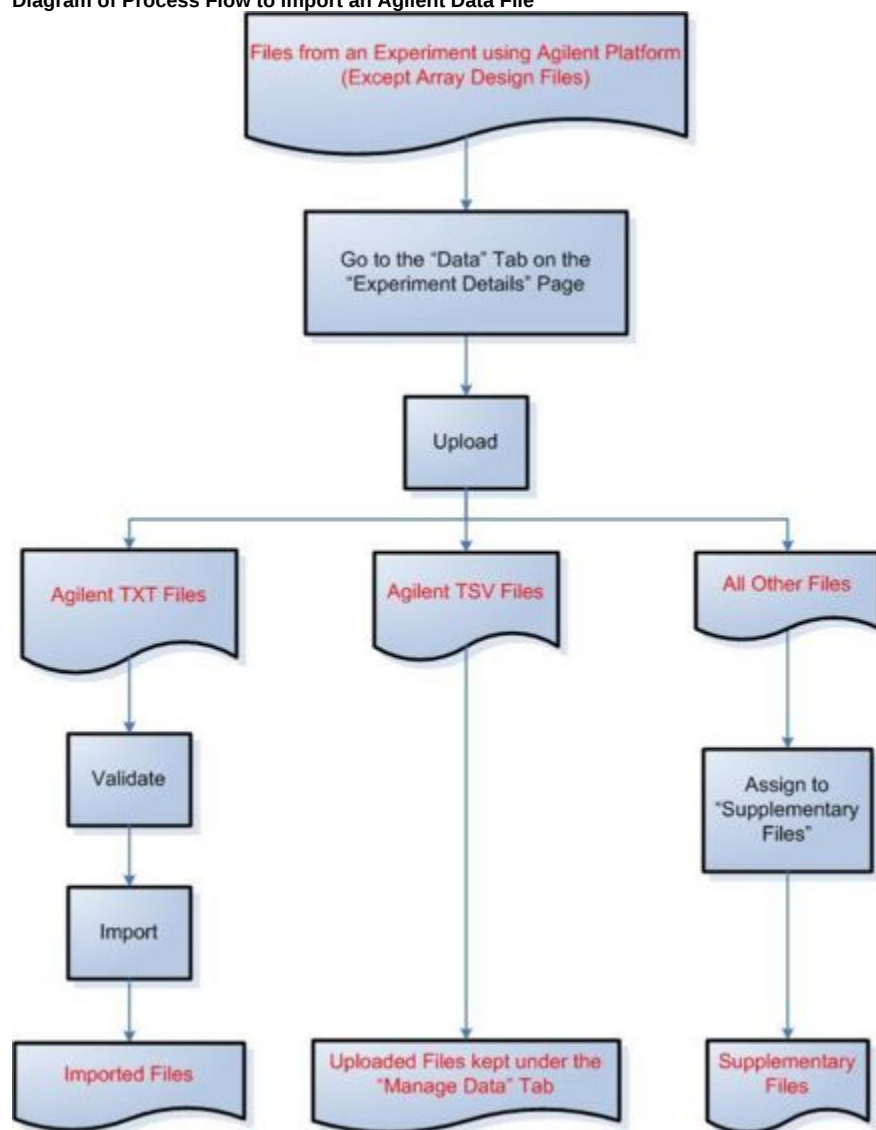


Figure 8. Import experimental data using the Agilent platform.

According to the figure above, to import files containing experimental data from the Agilent platform, first select the **Data** tab on the **Experiment Details** page, then proceed to upload the file. For a TXT file, you should validate the file, then import it. TSV files should not be validated or imported. They should instead be kept under the Manage Data tab. You should assign all other file formats to the **Supplementary Files** category.

Step 3F Import experimental data using the ImaGene platform

Because of a known issue concerning importing ImaGene TIF files, uploaded ImaGene TIF files should remain in the "Manage Data" tab under the "Data" tab for a selected experiment for caArray versions up to version 2.1.1. Once the issue is resolved, ImaGene TIF files will be able to be imported the same way as TXT files.

Diagram of Process Flow to Import an ImaGene Data File

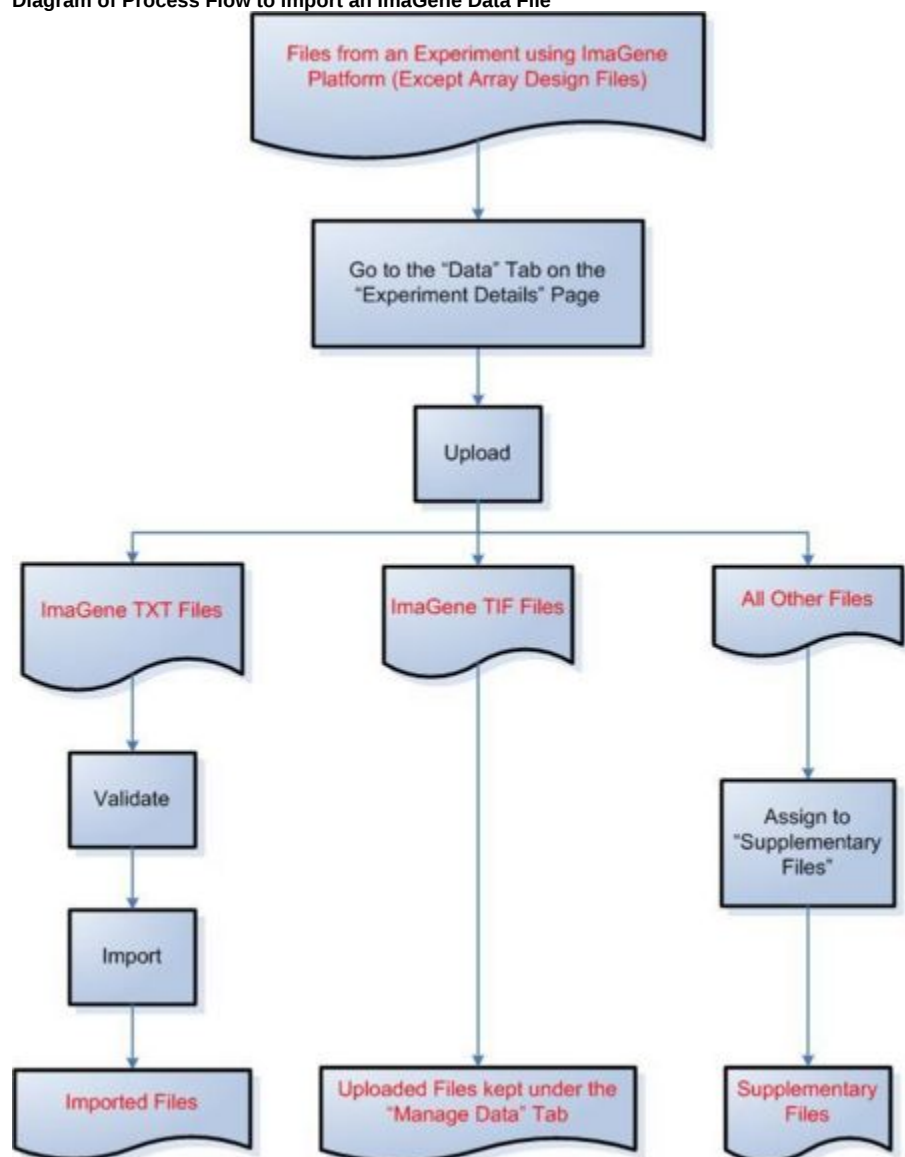


Figure 9. Import experimental data using the ImaGene platform.

According to the figure above, to import files containing experimental data from the ImaGene platform, first select the **Data** tab on the **Experiment Details** page, then proceed to upload the file. For a TXT file, you should validate the file, then import it. TIF files should not be validated or imported, and should instead be kept under the Manage Data tab. You should assign all other file formats to the **Supplementary Files** category.

Step 3G Import experimental data using the Nimblegen platform

Diagram of Process Flow to Import a Nimblegen Data File

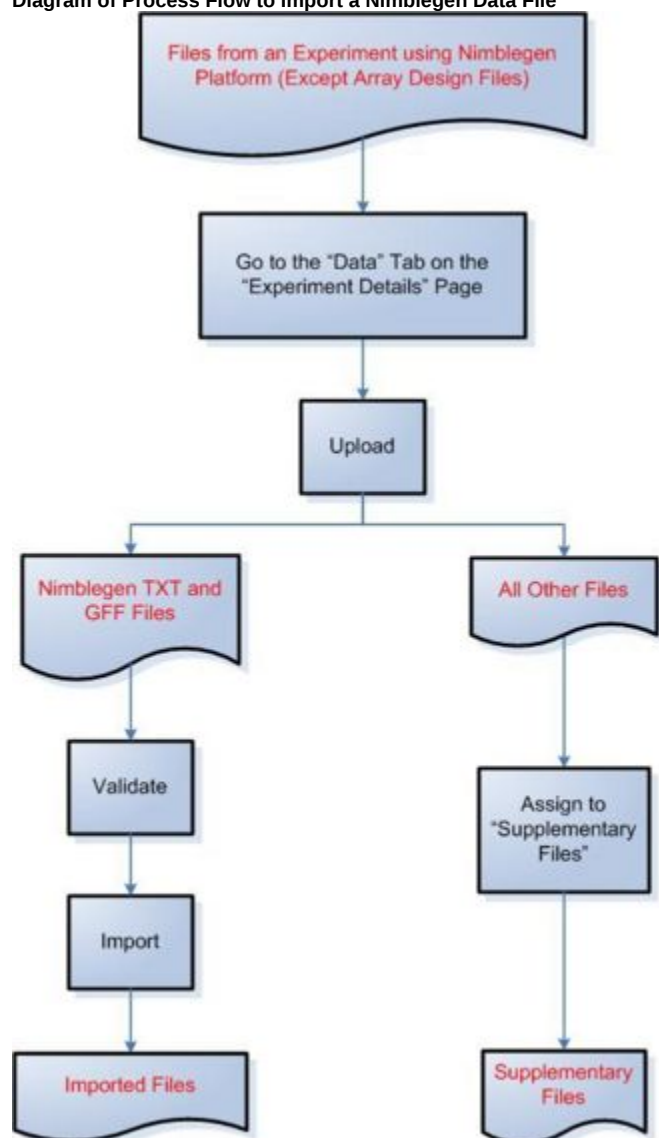


Figure 10. Import experimental data using the Nimblegen platform.

According to the figure above, to import files containing experimental data from the Nimblegen platform, first select the **Data** tab on the **Experiment Details** page, then proceed to upload the file. For a TXT or GFF file, you should validate the file, then import it. You should assign all other file formats to the **Supplementary Files** category.

Have a comment?

Please leave your comment in the [caArray End User Forum](#).