

caArray 016 - File Parsing into caArray

Question: Which file can be parsed into caArray? What is the benefit of file parsing?

Topic: caArray Usage

Release: caArray 2.2.0 and above

Date entered: 03/30/2009

Answer

Files Recognized by caArray

caArray has the ability to upload the array design file or experiment data from many array providers, even if it doesn't have a parser available yet. Those files will be imported into caArray without being validated or parsed. Even if a file is not parsed, the user will still be able to download the file (through the user interface as well as through the programmatic API), and will be able to associate the file to samples, extracts, and hybridizations. This feature allows data to be shared and help the system identify which new parsers are need developed in the future. For more information on how files are processed by caArray, review [caArray 017 - Meaning of the caArray Status of Importing - Imported versus Imported Not Parsed](#).

The following table, File types from caArray, (from Chapter 7 in the [caArray User Guide](#)) summarizes the file types that caArray currently supports with full validation and parsing as well as those that can be imported without validation and parsing. The user's guide also summarizes the [array design file types](#).

Files That Can Be Imported into caArray

File Types	Imported after validation and parsing	Imported without validation and parsing
Raw /processed data files; provide numerical values of array data	<ul style="list-style-type: none">Affymetrix CEL, CHP, CNCHPGenePix GPR*Illumina CSV, Sample Probe Profile TXT, Genotyping processed data matrix TXT, Raw TXTAgilent Raw TXTNimblegen Normalized Pair Report TXT <p>*For more information about GenePix GPR and Illumina CSV files, see MAGE-TAB SDRF Validation Rules, items 1 and 2.</p>	<ul style="list-style-type: none">Affymetrix DAT, RPT, TXT, and EXPAgilent TSV, derived TXTIllumina IDAT, TXTImaGene TIF, TXTNimblegen GFF, Raw or Derived TXTScanArray CSVGEO SOFT, GSM <p>Note: caArray may have new parsers available for data files in the system that are already imported but not parsed. To learn about retrofitting those files, see #Retrofitting Data Files.</p>
Array Design files; provide the design of an array.	For information about array design file types, see the table in Managing Array Designs .	
MAGE-TAB files; used to annotate experiments automatically	<ul style="list-style-type: none">MAGE-TAB SDRF (Sample and Data Relationship Format)MAGE-TAB IDF (Investigation Description Format) only, no referenced SDRFsMAGE-TAB Copy Number Data Matrix <p>Note: Only one IDF is allowed per import, since the import is in the context of a single experiment.</p>	<ul style="list-style-type: none">MAGE-TAB ADFMAGE-TAB Data Matrix (not copy number)
Supplemental Files	These cannot be validated nor imported. Files of unknown file type or simply reference files fall under this category. For more information, see #Supplemental Files .	



Note

Image files cannot be validated or imported successfully into caArray 2.4. For more information, see [Image File Importing Issues in caArray](#) as well as [Appendix B - Importing Data Files](#).

Benefit of File Parsing

For the data that are parsed into caArray, an analytical service (like [geWorkbench](#)) can pull the data out using the programmatic API and perform analysis on it or plot graphs from it etc. Another example is [web Genome](#), a caArray client, which pulls parsed data from caArray experiments and plots log ratio values against the chromosome location. With parsed data, a client can ask for quantitative types (columns) of data of interest, instead of having to retrieve the entire contents of the data file.

Have a comment?

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