CaArray2.3.0Importer Documentation

Description: A module to import data from caArray into GenePattern.

Author: Jared Nedzel, gp-help@broadinstitute.org

Summary: The CaArray2.3.0Importer imports data files from a caArray 2.3.0 repository into GenePattern. The module connects to a caArray 2.3.0 repository and then retrieves all files of a given extension for a named experiment. The retrieved files are collected into a single ZIP file archive which is returned as module output. Typical use of this module would be to retrieve all '.cel' files from an experiment in caArray and to subsequently pass the resulting zip file to the ExpressionFileCreator module for processing into a GenePattern gct or res file format for subsequent analysis.

This module is not compatible with earlier caArray versions.

Parameters:

Name	Description
url	URL to the caArray 2.3.0 repository. The default is to use the public caArray instance hosted at the NCI (https://array.nci.nih.gov:8080)
experiment	Title or public identifier of the experiment in caArray that data is to be imported from.
type	Type of bioassay data to be retrieved (RAW or DERIVED).
extension	Optional argument. Data file extension to be retrieved. Only data files with this extension will be imported into GenePattern. If this is not specified, all files of the specified type (RAW or DERIVED) will be retrieved.
zipFileName	Output file name that will be used for the retrieved data. Default is to use the experiment name with any spaces replaced by underscores.
username	Username for the caArray server. This is only required if you are accessing private data.
password	Password for the caArray server. This is only required if you are accessing private data.

Output Files:

1. a ZIP file containing all files of the given extension from the named experiment

Platform dependencies:

Task type: caBig
CPU type: any

OS:	any
Java JVM level:	1.5
Language:	Java