

Tutorial using the new Minid and BDBag systems with PeptideAtlas and the Trans-Proteomic Pipeline

Updated: 2017-10-31 - Eric Deutsch

If you have questions or suggestions with this tutorial, please contact edeutsch@systemsbiology.org

0) Before you begin

This tutorial assumes that you already have *Trans-Proteomic Pipeline* (TPP) 5.1.0 or later installed on your Microsoft Windows system (tested with Windows 7 Pro 64-bit). If that is not the case, install that first by downloading and following the instructions at

http://tools.proteomecenter.org/wiki/index.php?title=TPP:5.1_Installation. If you have a different kind of operating system, this tutorial should all still work, but the details of the installation will be somewhat different.

Make sure that you are connected to the Internet for this tutorial!

1) Test the FetchDataset system and check dependencies

- Open the *Petunia* TPP Graphical User Interface (TPP GUI) and login
- Under the [TPP Tools] menu, click [Fetch Datasets] near the bottom of the [Utilities] column
- Click on [Show version information and available features]
- Check to see if *Python*, *Minids*, and *BDBags* are all supported and available. If you already have these packages on your system, then you can proceed to step 5
- If not, install them as described in steps 2-4 below

2) Install Python

- Any version of Python 2 or 3 is probably fine, but we recommend and assume Python **3.6.x**
- Note that it is important how you install Python. If you only install it for your user, then the *Petunia* interface will not be able to see it and use it. It must be installed for all users.
- For Windows, go to <https://www.python.org/downloads/> and download python 3.6.x (as of this writing, click on the large yellow button [Download Python 3.6.3] although that may change.)
- Run the downloaded Python installer *python-3.6.x.exe*
- Check the option “Add Python 3.6 to PATH” at the very bottom of the first window
- Then click the [Customize Installation] option
 - All options should be checked, click [Next]
- Check option [Install for all users]. This may require administrator password but is required
- Click [Install]
- After installation there may be an option to increase the command line length beyond MAX_LIMIT. We encourage you to perform this extra action, although it is not necessary for this tutorial.
- If there are installation errors, resolve these before proceeding.
- Make sure you can type this on the command-line (to open a cmd window, click the Windows button, and type **cmd** in the search box and hit [ENTER])
 - `python --version`
and get a result similar to: *python 3.6.3*
- You will need to reboot or restart the *TPP Apache Web Service* for the changes to PATH to be applied to the web server, which the TPP GUI system uses. If you don't do this, TPP GUI won't find *Python*.

3) Install Minid software

- Go to: <https://github.com/ini-bdds/minid>
- Click the green [**Clone or download**] button
- Choose [**Download ZIP**] and save file
- Open the *Downloads* folder, right-click on the *minid-master.zip* file and choose [**Extract All...**]
- Open a *cmd* window running as Administrator and install the python module.
 - To open a cmd window as Administrator, click the Windows button, and type **cmd** in the search box (but don't hit [Enter]). When cmd.exe is found and listed in the search results, **right click** and select '**Run as Administrator**'. This will allow you to write to the system Python library directories:
 - `cd \Users\yourname\Downloads\minid-master\minid-master` (or wherever the extract appeared)
 - `python setup.py install`
- Review the installation message and if there are any errors, resolve before proceeding

4) Install BDBag software

- Go to: <https://github.com/ini-bdds/bdbag>
- Click the green [**Clone or download**] button
- Choose [**Download ZIP**] and save file
- Open the *Downloads* folder, right-click on the *bdbag-master.zip* file and choose [**Extract All...**]
- Open a *cmd* window running as Administrator (as above) and install the python module:
 - `cd \Users\yourname\Downloads\bdbag-master\bdbag-master` (or wherever extract appeared)
 - `python setup.py install`
- Review the installation message and if there are any errors, resolve before proceeding

5) Check the fetchDataset system and dependencies

- Open the *Petunia* interface and login
- Under [TPP Tools] menu, select [**Fetch Datasets**] near the bottom of the [Utilities] column
- Click on [**Show version information and available features**]
- You should see that *Python*, *Minids*, and *BDBags* are all supported and available
- If not, resolve any errors before proceeding or retry the above steps
- You should also see that the *fetchDataset* version installed matches the latest available
- If not (i.e. your version of *fetchDataset* is older than the latest), you should update to the latest before proceeding by clicking on the [**Update to the latest version of fetchDataset**] link at the bottom of the INFO box.
- When all your software and dependencies are up-to-date, then processed to step 6

6) Find a dataset

- Go to the Human Plasma PeptideAtlas data repository at: <http://www.peptideatlas.org/hupo/hppp/repository/>
- Find the data from the recent Vogt et al. article. In the Search box at the top, type in "**vogt**" and [Enter]

- The HCD data are probably better, so try the first one, the *PAe005719*
- Expand it by clicking on **[+]** next to *PAe005719*
- Notice that the data are available via *Minids* for this dataset. Let's choose to download just the end product, the "**Proteins**" *BDBag* which has an **ark:/57799/b96q2x**
- Click the little clipboard icon next to **ark:/57799/b96q2x** to copy it to the Windows clipboard
- Notice the accessible information for *Minids* and *BDBags*. Click on "What's a Minid?"

7) Download the dataset in TPP

- Open the *Petunia* interface and login
- First, let's create a place to put the dataset
 - Click on **[Files]** from the main menu
 - In the lower right text box right below **[Create a new directory]**, type in "**PublicDatasets**" (no spaces and without the quotation marks), and then click on **[Create a new directory]**
 - Click on the newly created folder **PublicDatasets**
 - In the lower right text box right below **[Create a new directory]**, type in "**PAe005719**" (no spaces and without the quotation marks), and then click on **[Create a new directory]**
- Now let's download the data to there
 - Under **[TPP Tools]** menu, select **[Fetch Datasets]** near the bottom of the **[Utilities]** column
 - Choose a target directory by clicking **[Add Files]**
 - Navigate into **PublicDatasets** and check the **PAe005719** item and click **[Select]**
 - In the SPECIFY DATASET ID section, paste in the *Minid*: **ark:/57799/b96q2x** (this *Minid* should be in your clipboard from step 6, so just paste with *CTRL+v*)
 - Check the option to **purge all contents** of the destination folder
 - This "final results" *BDBag* is quite small at only 74 MB in order to make this a speedy tutorial. Selecting other *BDBags* may incur substantially longer download times.
 - Click **[Fetch Dataset]**
- Let the process run and examine the output carefully for any errors
- If there are errors, resolve before proceeding

8) Examine the dataset

- Click on **[Files]** from the main menu
- Navigate to **PublicDatasets/PAe005719** folder
- Click on the **[ProtXML]** link to view the proteins detected
 - See protein list
 - Click on the **21** in the first column, next to **group #2**. See the mapping of peptides across the members of the group
 - Click on a protein **NX_P35579-2** to explore peptide mappings for isoform 2 of Human Myosin-9
 - The dataset may be explored further as desired

Finis