# Overview

1. pQuant is written in Java, so it could run in Windows or Linux/Unix operating system.

2. pQuant is designed to read the identification results from pFind/pBuild or ProLuCID/DTASelect, and quantifies the peptides and proteins.

3. We are developing the workflow of pFind/pBuild/pQuant, which is named pFindStudio (v3.0). pFindStudio (v3.0) will be installed in Windows with the elaborate UI.

4. Also, pQuant could run in Linux/Unix with ProLuCID/DTASelect, just like Census.

5. This manual describes how to operate pQuant with MS1 files and DTASelect-filter.txt file in Windows. It is a little difficult to prepare a pQuant manual for Linux/Unix, because these operating systems are different at different clusters. However, an expert could easily operate pQuant in Linux/Unix according to this manual.

# List of pQuant files

1. File “pQuant.jar”. This jar file is pQuant. It could run in Windows or Linux/Unix without any revision. It was built under JRE 1.7.

2. Folder “ini”. Three files are in this folder: “aa.ini”, “element.ini”, and “modification.ini”. These files provide the basic information. Generally, they don’t need to be revised.

3. File “pQuant\_cfg.txt”. All parameters are set in this file. This file needs to be revised before launching pQuant. You can double-click pQuant.jar to get a new pQuant\_cfg.txt.

4. File “run\_pQuant.bat”. Double-click to launch pQuant. This file needs to be revised before the double-clicking. Open it with notepad in Windows, then you will know how to launch pQuant in Linux/Unix.

# Launch pQuant

1. The “pQuant.jar” could be placed in any directory of your system. For example, it is placed in “D:\pQuant\”. Also, the “pQuant.jar” could be placed in Linux/Unix.

2. Double-click pQuant.jar means launch pQuant without arguments. In this case, pQuant generates a new task folder; there is a new pQuant\_cfg.txt in it.

3. Revise the run\_pQuant.bat to launch pQuant with the only argument, i.e., pQuant\_cfg.txt.

4. Before launching pQuant, the following works should be done.

# Get MS1 files and DTASelect-filter.txt

1. [RawXtract]. Export Thermo RAW files to MS1 and MS2 files.

2. [ProLuCID]. Search the MS2 files, getting the SQT files.

3. [DTASelect]. Filter the SQT files, getting the DTASelect-filter.txt file.

4. Download the MS1 files and DTASelect-filter.txt from the cluster to your PC. For example, these files are downloaded to the folder of “D:\pQuant\data-test\”. Also, you can keep these files in the cluster, and launch pQuant in Linux/Unix.

# Revise ini files

1. We suggest placing the “ini” folder in “D:\pQuant\ini\”.

2. “element.ini”. The item “EB105=15N|14.0030732,15.0001088,|0.01,0.99,|” means the 15N labeling efficiency is 99%.

3. “aa.ini”. The residue C should be “R3=C|C(5)H(8)N(1)O(2)S(1)14N(1)|” if Carbamidomethyl is considered as a fixed modification in DTASelect-filet.txt. The element of 14N will not be changed to 15N in quantitation. It is very important. Without this revision, pQuant will get incorrect masses and incorrect ratios for the peptides with the residue C.

4. “modification.ini”. It’s used in complicated PTM-identification and PTM-quantitation.

# Revise pQuant\_cfg.txt file

1. We suggest placing the pQuant\_cfg.txt in Folder “D:\pQuant\Task\_201406101941\”.

2. Open the file, and revise the parameters one by one.

3. We explain some parameters here. Generally, just basic parameters are need to be set.

//-------------------------

PATH\_INI\_ELEMENT

The path of element.ini. PATH\_INI\_ELEMENT=D:\pQuant\element.ini;

//-------------------------

“PATH\_INI\_MODIFICATION” and “PATH\_INI\_RESIDUE” are similar.

//-------------------------

“PATH\_BIN”, “NUMBER\_THREAD” and “TYPE\_STRAT” are used in pFindStudio to process large scale datasets.

//-------------------------

PATH\_MS1

Folders or files are all legal. Use | as separator when multiple files or folders are set. For example, PATH\_MS1=D:\Bio\_1\_MS\_1\| D:\Bio\_1\_MS\_2\| D:\Bio\_1\_MS\_3\|;

//-------------------------

EXTENSION\_TEXT\_MS1

pQuant will process all files with this extension in one folder. It is case sensitive.

//-------------------------

PATH\_IDENTIFICATION\_FILE

Use | as separator when multiple files are set.

For example, PATH\_IDENTIFICATION\_FILE= D:\Bio\_1\_MS\_1\DTASelect-filter.txt| D:\Bio\_1\_MS\_2\ DTASelect-filter.txt | D:\Bio\_1\_MS\_3\ DTASelect-filter.txt |;

//-------------------------

TYPE\_IDENTIFICATION\_FILE

For DTASelect-filter.txt, TYPE\_IDENTIFICATION\_FILE=1; for pFind.spectra, TYPE\_IDENTIFICATION\_FILE=2;

//-------------------------

THRESHOLD\_FDR

Used for pFind3.0. “THRESHOLD\_FDR=0.01;” means get all PSMs with FDR less than 0.01;

//-------------------------

TYPE\_LABEL

“TYPE\_LABEL=0;” means it is a labeling quantitation; “TYPE\_LABEL=1;” mean it is a label free quantitation.

//-------------------------

LL\_INFO\_LABEL

“LL\_INFO\_LABEL=2|none|R:\*{N,15N}|;” means 15N-labeling. The item “2” means two isoforms involved. “none” means nothing changed. “R” is Residue. “\*” means all resides. “<N,15N>” means the element of N will be changed to 15N in quantitation.

For SILAC, LL\_INFO\_LABEL=2|none|R:K{N,15N}R:K{C,13C}R:R{N,15N}R:R{C,13C}|;

//-------------------------

LL\_FLAG\_ENRICHMENT\_CALIBRIATION\_15N

It is not 15N-labeling, LL\_FLAG\_ENRICHMENT\_CALIBRIATION\_15N=0; it is 15N-labeling but do not calibrate the labeling efficacy, LL\_FLAG\_ENRICHMENT\_CALIBRIATION\_15N =1; it is 15N-labeling and calibrate the labeling efficiency, LL\_FLAG\_ENRICHMENT\_CALIBRIATION\_15N =2;

If LL\_FLAG\_ENRICHMENT\_CALIBRIATION\_15N =2, pQuant will calibrate the labeling efficiency for each peptide. We recommend that use it for checking the labeling efficiency. If there are interferences, the calculated labeling efficiency is not reliable and the ratios are inaccurate.

//-------------------------

NUMBER\_SCANS\_HALF\_CMTG

NUMBER\_SCANS\_HALF\_CMTG=200 is appropriate. If the scan of MS/MS is 10000, pQuant will check the signals of full MS scans 9800 - 10200.

//-------------------------

PPM\_FOR\_CALIBRATION

If the mass spec is calibrated, it is 0.

//-------------------------

PPM\_HALF\_WIN\_ACCURACY\_PEAK

Similar to precursor mass tolerance in identification. “PPM\_HALF\_WIN\_ACCURACY\_PEAK=10;” means the deviations between the masses of experimental peaks and theoretical peaks are less than 10 ppm.

//-------------------------

NUMBER\_HOLE\_IN\_CMTG

It is used when pQuant determines the start and end point of a chromatogram. “NUMBER\_HOLE\_IN\_CMTG=2;” means terminals are found when two holes (extremely low signals) appears in a chromatogram.

//-------------------------

TYPE\_SAME\_START\_END\_BETWEEN\_EVIDENCE

Use “TYPE\_SAME\_START\_END\_BETWEEN\_EVIDENCE=0;” in analyzing the 1:1 mixed labeling samples.

“TYPE\_SAME\_START\_END\_BETWEEN\_EVIDENCE=1;” and “TYPE\_SAME\_START\_END\_BETWEEN\_EVIDENCE=2;” are used in label free.

//-------------------------

TYPE\_PROTEIN\_RATIO\_CALCULATION

“TYPE\_PROTEIN\_RATIO\_CALCULATION=0;” means get median of the peptide ratios as the protein ratio.

“TYPE\_PROTEIN\_RATIO\_CALCULATION=1;” means use kernel estimation method to get the protein ratio.

“TYPE\_PROTEIN\_RATIO\_CALCULATION=2;” means report the two above mentioned ratios.

//-------------------------

TYPE\_UNIQUE\_PEPTIDE\_ONLY

“TYPE\_UNIQUE\_PEPTIDE\_ONLY=0;” means all peptide ratios are used in calculate the protein ratios.

“TYPE\_UNIQUE\_PEPTIDE\_ONLY=1;” means only unique peptide ratios are used in calculate the protein ratios.

//-------------------------

THRESHOLD\_SIGMA

“THRESHOLD\_SIGMA=0.5;” means only the peptide ratios with the “sigma”s less then 0.5 are used in calculate the protein ratios.

//-------------------------

DIR\_EXPORT

The path for the results.

//-------------------------

TYPE\_DIR\_EXPORT

“TYPE\_DIR\_EXPORT=1;” means a new sub-folder with time-marks will be created. It is used when testing different parameters, because new results will not cover the former ones.

“TYPE\_DIR\_EXPORT=0;” means results will be exported in this folder.

//------------------------

# Revise run\_pQuant.bat file

1. We suggest placing the file in Folder “D:\pQuant\”.

2. There is only one command in this file, i.e. “launch pQuant with the pQuant\_cfg.txt”.

//------------------------

java -jar -Xmx4000m pQuant.jar D:\pQuant\Task\_201406101941\pQuant\_cfg.txt

pause

//------------------------

3. Change pQuant\_cfg.txt for different tasks.

# Check the results

1. The results could be found in “DIR\_EXPORT” you set. These result files are texts, which could be easily read by other programs or scripts.

2. There are four files in the result folder.

//------------------------

pQuant.spectra

“pQunat.spectra” provides details for each quantitation result in PSM level. It is a text file, and could be open by Notepad or Ultraedit, etc.

“@” means a new quantitation result.

I,Q,01: The number of evidences involved in this quantitation.

I,Q,02: Labeling efficiency. It is a valid value when LL\_FLAG\_ENRICHMENT\_CALIBRIATION\_15N=2 in 15N-labeling. Otherwise, it is -0.0716.

I,Q,03: Ratio and sigma calculated based on mono isotopic peaks. Ratio=Evidence1/Evidence0.

I,Q,04: Ratio and sigma calculated based on isotopic peaks with the least interference. We report this ratio in pQuant paper and in pQuant.spectra.list.

I,Q,05: Ratio and sigma for testing. Users could ignore this information.

I,E,0,01: The identification information of 0-index evidence, they are “Sequence, Modification, Score, ProteinID”.

I,E,0,02: Abundances of theoretical distribution of this peptide.

I,E,0,03: Masses of theoretical distribution of this peptide.

I,E,0,04: List of full MS1 scans.

I,E,0,05: The chromatograms. The rows correspond to theoretical peaks (I,E,0,02 and I,E,0,03), and the column correspond to list of full MS1 scans (I,E,0,04).

I,E,0,06: The index of monoisotopic peak, the index of isotopic peak with the least interference.

I,E,0,07: The start and end points of the chromatograms. This signals are used to calculated the ratios. In labeling quantitation, “I,E,0,07”, “I,E,0,07”, “I,E,1,07” and “I,E,1,08” are almost the same.

I,E,0,08: The start and end points of the chromatograms. In label free quantitation, evidences come from different raw files, so they have different start and end points.

I,E,0,02~ I,E,0,08 could be used to plot the chromatograms, just like we did in pQuant paper.

I,E,1,01~I,E,1,08: Another evidence.

//------------------------

“pQuant.spectra.list” is an excel-format text file. It could be open by excel for further analysis.

Column 1: Spectrum ID.

Column 2: Sequence

Column 3: Modification

Column 4: Score (read from identification result)

Column 5: Total Intensity read from DTASelect-filter.txt. For other files, it is 0.0.

Column 6: AC or protein ID.

Column 7: DE. It will be null if DEs are not provided by identification results.

Column 8: Ratio 2 (I,Q,04).

Column 9: Sigma 2 (I,Q,04). If there are three evidences, there are two such ratios, they are Ratio=Evidence1/Evidence0 and Ratio=Evidence2/Evidence0.

Column 8 Column 9 Column 10 Column 11

Ratio2 Sigma2 Ratio2 Sigma2

Column 10: Ratio 1 (I,Q,03).

Column 11: Ratio 1 (I,Q,03). If there are three evidences, there are two such ratios, they are Ratio=Evidence1/Evidence0 and Ratio=Evidence2/Evidence0.

Column 10 Column 11 Column 12 Column 13

Ratio1 Sigma1 Ratio1 Sigma1

//------------------------

pQunat.proteins

“pQunat.proteins” provides details for each quantitation result in protein level. It is a text file, and could be open by Notepad or Ultraedit, etc.

“@” means a new quantitation result. AC DE.

Group ID. Group Type.

Ratio calculated by median methods.

Ratio calculated by kernel estimation method.

The number of peptide ratios calculated. The number of peptide ratios used in calculating protein ratios.

Spectrum ID, Sequence, Ratio2 (I,Q,04), Sigma2, Type\_Infer (0: used in calculating protein ratios).

//------------------------

“pQuant.proteins.list” is an excel-format text file. It could be open by excel for further analysis.

Column 1: Protein ID (AC)

Column 2: Ratio calculated by getting median method.

Column 3: Sigma calculated by getting median method.

Column 4: Ratio calculated by kernel estimation method.

Column 5: Sigma calculated by kernel estimation method.

Column 6: The number of peptide ratios calculated.

Column 7: The number of peptide ratios used in calculating protein ratios.

Column 8: Information of all peptide ratios (column6). Ratio,Sigma,Type\_Infer (0: used in calculating protein ratios).

//------------------------

5. We think that special plotting or presentation scripts are prerequisite for publication. “pQuant.spectra” has provided enough information. Chao Liu writes some Matlab scripts to present the chromatograms in the pQuant paper.

# Further analysis

1. We write special scripts for further analyses.

2. Protein ratios are (re)calculated, because their peptides may come from different experiments, and different strategies (i.e. whether to discard the one-hit-wonders, whether to discard inaccurate peptide ratios) may be involved. Kernel estimation method is used in this step.

3. Normalize the protein ratios.

4. Significance analysis. We use the statistical method to get significantly changed ratios, and use the fold-change method to get “just” changed ratios. Spectra-counting method is also used when some protein is not identified in some experiment.

5. Above is the data processing of pQuant. In the future, we may set foot on bioinformatics.