

SAVControl(v1.0) User's Manual

1 Overview

The SAVControl software is developed for site-level quality control of variant peptide identifications. It first filters out false variant peptide identifications by Transfer FDR control, and then evaluates the reliability of the variant sites by unrestricted delta mass relocation (PTMiner) and introducing alternative interpretations, e.g. modifications. All variant sites are finally classified into three levels: Level I (reliable results), Level II (ambiguous results) and Level III (unreliable results). SAVControl was written in standard C# language on platform of Microsoft Visual Studio, and has a user-friendly interface. It can currently process database search results of the Myrimatch and Mascot search engines. SAVControl is freely available at <http://fugroup.amss.ac.cn/software/SAVControl/SAVControl.html>.

2 Running SAVControl

Now SAVControl can be run on Windows system smoothly, requiring the MATLAB Runtime (MCRInstaller.exe (version R2014a (8.3))) to be installed in advance. The MATLAB Runtime can be downloaded from <https://cn.mathworks.com/products/compiler/mcr.html>. After completing the preparation work, the user can double-click "SAVControl.exe" and set parameters to run SAVControl (Figure 1).

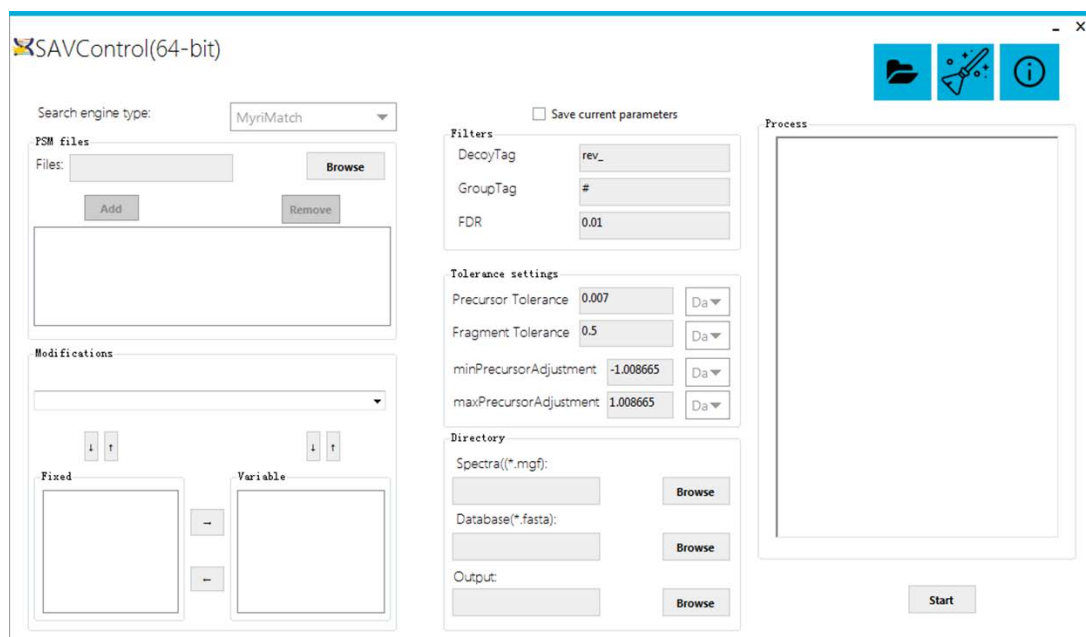



Figure 1. The interface of SAVControl.

2.1 Setting parameters and generating a new parameter file in GUI

2.1.1 Search engine type (Figure 2a). The SAVControl software can currently process database search results of the Myrimatch and Mascot search engines.

2.1.2 PSM files (Figure 2b). SAVControl supports following types of file: DAT file (Mascot (v2.5.1)), pepXML file (MyriMatch (v1.5.6)). Click “Browse” button to select the PSM files. Click “Add” button to add the selected PSMs. Click “remove” to remove the selected PSMs. Before the database searching, it is worth noting that the searched database is the target-decoy database, which is composed of forward protein sequences and their reversed ones. In particular, do not use the decoy database made by mascot search engine. Set the FDR threshold in the search engine as 100% to get all the peptide spectrum matches (PSMs).

2.1.3 Modifications (Figure 2c). To specify fixed/variable modifications for SAVControl, all the modifications used in database search must be set. A) Fixed modification(s). Select the types of fixed modifications used in database search. B) Variable modification(s). Select the types of variable modifications used in database search. SAVControl lists all modifications in Unimod database.  button can be

used to select or delete modification(s). or button can be used to exchange modification(s) between fixed modification(s) and variable modification(s).

2.1.4 Filters (Figure 2d). A) DecoyTag. The decoy tag is used to label the decoy protein. Default is “rev_”. B) GroupTag. The group tag is used to label the group protein. Default is “#”. C) FDR. Set the threshold of the false discovery rate. Default is 0.01.

2.1.5 Tolerance settings (Figure 2e). A) Precursor tolerance. This parameter is used to specify the instrument mass precision/tolerance of peptide precursors. B) Fragment tolerance. This parameter is used to specify the instrument mass precision/tolerance of fragment ions. C) minPrecursorAdjustment. This parameter is the minimum step size for calibrating the mass error of the parent ion by MyriMatch search engine. D) maxPrecursorAdjustment. This parameter is the maximum step size for calibrating the mass error of the parent ion by MyriMatch search engine.

2.1.6 Directory (Figure 2f). A) Spectra (*.mgf). Choose the location where the Mascot generic format (mgf) files is located by clicking the “Browse” button. B) Database (*.fasta). Choose the location where the protein database file is located by clicking the “Browse” button. C) Output. Choose the directory where the SAVControl results will be saved by clicking the “Browse” button.

2.1.7 Process (Figure 2g). When the background program is running, the process textbox at the right of the window would show the process of SAVControl. It includes transfer FDR control, unrestricted mass shift relocation and variant site assessment (Figure 3).

2.1.8 Save current parameters (Figure 2h). Saving current parameters is optional.

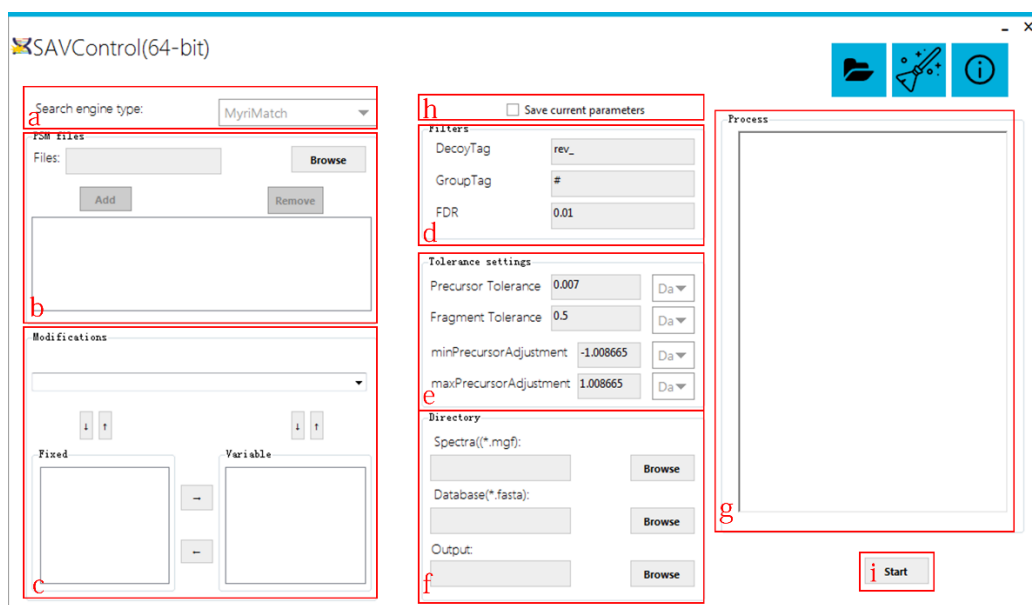


Figure 2. Search parameters setting.

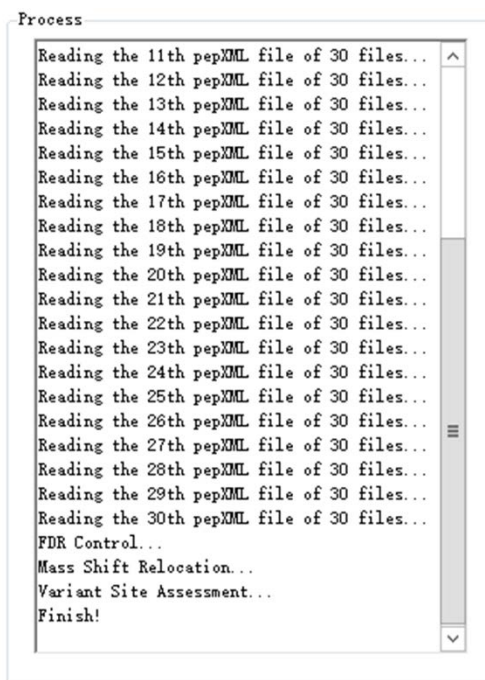


Figure 3. The process of SAVControl.

2.2 Loading an existing parameter file

If an existing parameter file (*.param) is available, a convenient way to set parameters is loading this parameter file directly (Figure 4a).

2.3 Clearing current parameters

A convenient way to clear the current parameters is clicking the clearing button (Figure 4b).

2.4 About SAVControl

The information of SAVControl can be obtained by clicking the button at Figure 4c (Figure 5).

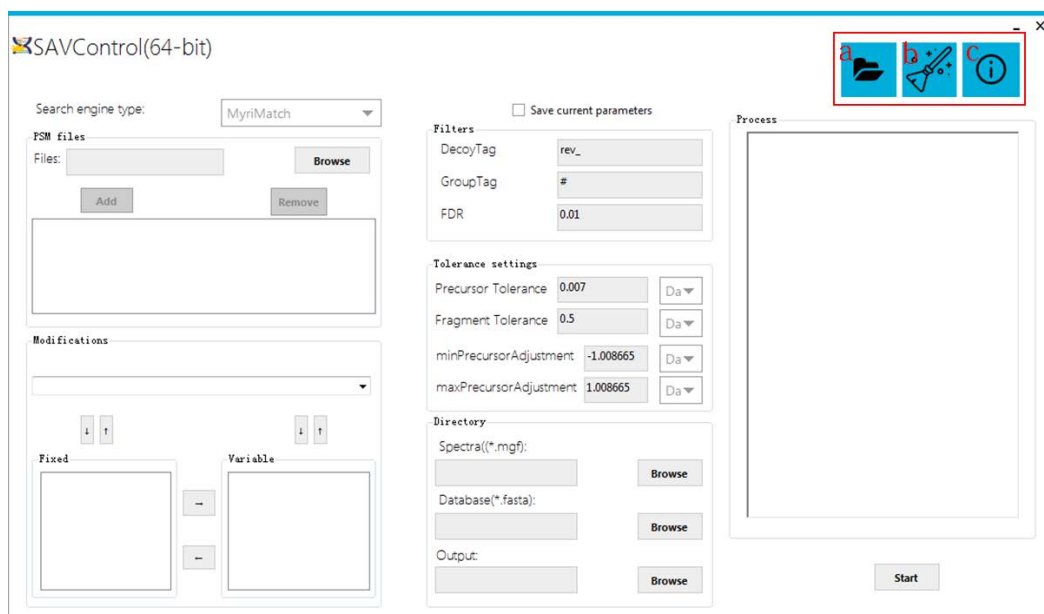


Figure 4. Interface of manually loading a parameter file, clearing current parameter, loading this manual file.

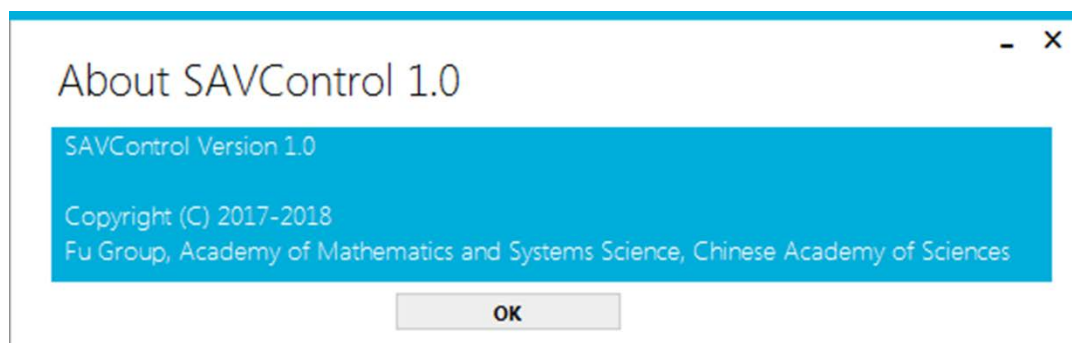


Figure 5. About SAVControl.

2.8 Running SAVControl

Make sure that all the parameters are appropriate, and then click the “Start” button to start SAVControl (Figure 2i). If the background program is done successfully, a dialog showing “SAVControl finished successfully!” would appear (Figure 6) and the process (Figure 2g) would show “Finish!”

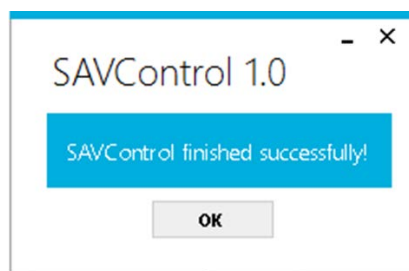


Figure 6. Dialog when the background program finished successfully.

3 Output File Formats

Once the calculation is done, SAVControl generates one file in the result directory (Table 1).

Table 1: the SAVControl result file.

File name	Annotation
SAVControlResult.txt	The SAVControl result in the folder specified in "SAVControlResult".

4 About SAVControlResult.txt

This file contains the information of the SAVControl results. The detailed description of every column in this file is given in Table 2.

Table 2: Descriptions of headers in SAVControlResult.txt.

Column Name	Description
DatasetName	Spectrum file name
Spectrum	Spectrum name
Protein	Identified protein name
OriginalSequence	Non-variant peptide sequence corresponding to the Identified variant peptide sequence
VariantSequence	Identified peptide sequence
BeforeModification	Non-variant peptide modification names and locations including both fixed and variable modifications. If there are more than one modifications, they are separated by “;”, e.g. “3,Oxidation (M);8,Carbamidomethyl (C)”. If there are no modifications, leave it as “-”.
AfterModification	Variant peptide modification names and locations including both fixed and variable modifications. If there are more than one modifications, they are separated by “;”, e.g. “3,Oxidation (M);8,Carbamidomethyl (C)”. If

	there are no modifications, leave it as “-”.
Score	The value of identified score by the search engine.
PTMinerSite	The position of relocation site(s) by PTMiner.
AA	The amino acid(s) corresponding to the relocation site(s) by PTMiner.
VariantSite	The variant site corresponding to the variant protein.
VariantInformation	The variant type corresponding to the variant protein.
Probability	The relocation site protein by PTMiner.
VariantDeltaMass	The variant deltamass corresponding to the variant type.
ObsMH	The experimentally observed peptide mass plus one proton.
TheoMH	The theoretical peptide mass plus one proton.
Charge	The charge of peptide precursor.
Explain	If the relocated site(s) can be interpreted by some entry in the Unimod database, the explain is these interpreted names. If not, the explain is “-”.
Level	Each identified site is classified into three levels: Level I, Level II and Level III.