InSyBio Datastore

InSyBio

Intelligent Systems Biology

User Manual

October 2018

Insybio Suite v2.4

www.insybio.com

INTRODUCTION

Data Store is a tool that helps you upload input files to the other three InSyBio tools (Interact, ncRNASeq and BioNets) and stores the datasets created by them.

USING INSYBIO DATA STORE YOU CAN:

- upload a file, and
- view, download or delete:
    - Your uploaded ncRNA Sequences Files
    - Your uploaded miRNA Sequences Files
    - Your uploaded mRNA Sequences Files
    - Your uploaded Biological Networks Files
    - Your uploaded Biological Network Clusters Files
    - Your uploaded SOFT Files
    - Your uploaded or saved Molecules Quantification Files
    - Your uploaded or saved Differential Expression Analysis Biomarkers Files
    - Your uploaded or saved Network Comparison Biomarkers Files
    - Your uploaded or saved Merged Biomarkers Files
    - Your saved PPI Datasets from InSyBio Interact Tool
    - Your uploaded or saved Biomarkers Datasets
    - Your uploaded or saved Biomarkers Labels Files
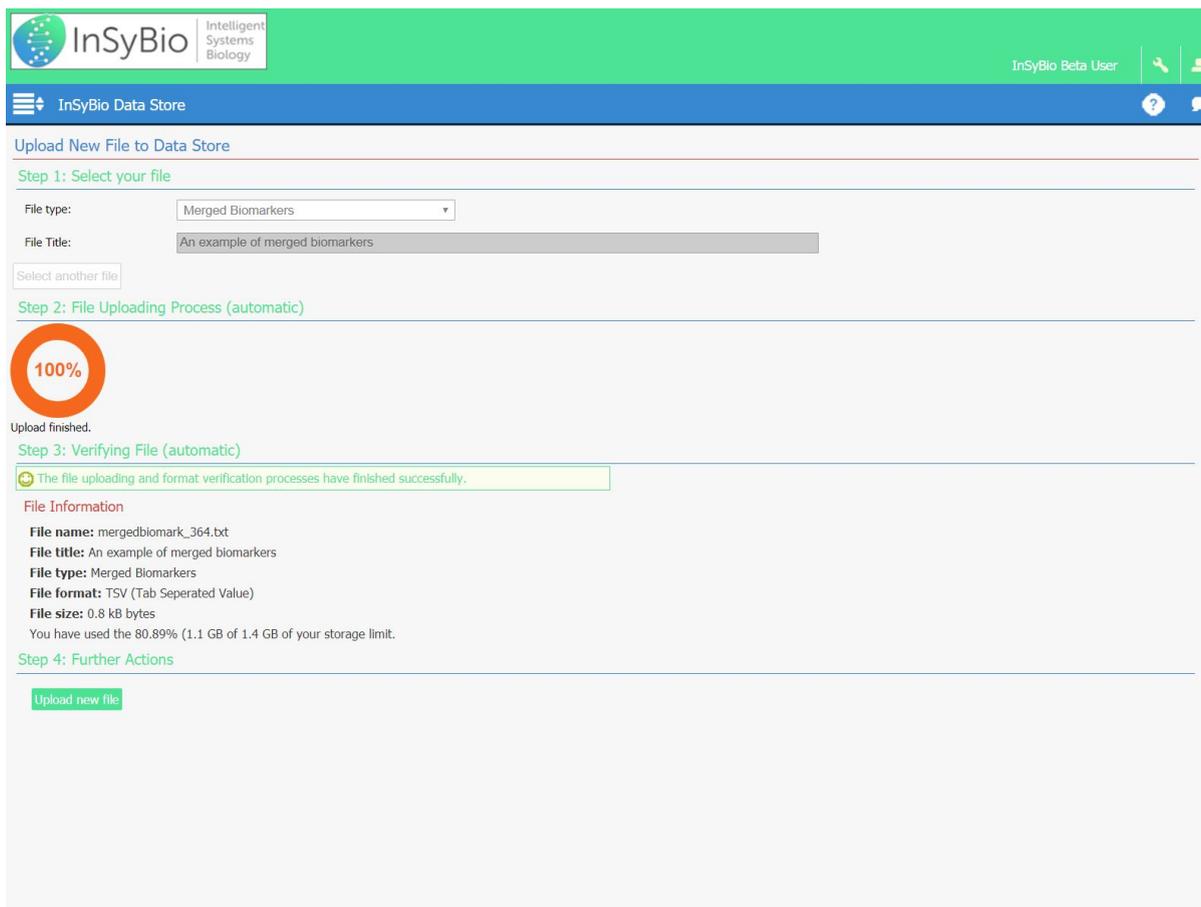   - Your saved Biomarkers Model Files

UPLOAD NEW FILE

In order to upload a new file you have to:

A. select one of the available file types (ncRNA Sequences, miRNA Sequences, mRNA Sequences, Biological Networks, Biological Network Clusters, SOFT Files, Molecules Quantification, Differential Expression Analysis Biomarkers, Network Comparison Biomarkers, Merged Biomarkers),
B. provide a title for your file,

C. select a file from your computer and

D. upload the selected file.

When the uploading process finishes, the file is verified automatically to check it has the right format.

After the verification process, the uploaded file information is presented (file name, file title, file type, file format, file size), together with a message informing you about your storage use. Your total storage limit depends on the edition of the InSyBio Suite you are using (Evaluation or commercial version).

The upload new file process ends with Data Store asking you to upload another file.

Selecting each of the available file types, you are redirected to a list of your respective datasets.

NCRNA SEQUENCES FILES

You can upload a new ncRNA Sequences File in fasta format with one or multiple sequences. The RNA sequences datasets which can be handled by the current version are those derived from RNA sequencing experimental techniques such as RNA-seq Next Generation sequencing.

You can find a list of your uploaded ncRNA sequences. In the list, you can view the file name, the file description, the type and the format of the file, the file size, the date of the file creation and the date of the last modification. You can always download or delete each file listed.

For an ncRNA sequences file type you can select one of the two available actions:

- Feature Calculation at ncRNASeq, or
- miRNA Prediction at ncRNASeq

and be guided to the respective functionality of ncRNASeq.

MIRNA SEQUENCES FILES

You can find a list of your uploaded miRNA sequences. In the list, you can view the file name, the file description, the type and the format of the file, the file size, the date of the file creation and the date of the last modification. You can always download or delete each file listed.

For a miRNA sequences file, you can select one of the two available actions:

- miRNA Target Feature Calculation at ncRNASeq, or
- miRNA Target Prediction at ncRNASeq

and after also selecting the matching mRNA sequences file, be guided to the respective functionality of ncRNASeq.

MRNA SEQUENCES FILES

You can find a list of your uploaded mRNA sequences. In the list, you can view the file name, the file description, the type and the format of the file, the file size, the date of the file creation and the date of the last modification. You can always download or delete each file listed.

For a mRNA sequences file, you can select one of the two available actions:

- miRNA Target Feature Calculation at ncRNASeq, or
- miRNA Target Prediction at ncRNASeq

and after also selecting the matching miRNA sequences file, be guided to the respective functionality of ncRNASeq.

BIOLOGICAL NETWORKS FILES

You can upload a new Biological Network File. The Biological Network datasets which can be handled by the current version are those derived from experimental techniques for the prediction of protein-protein interactions and protein complexes

such as Yeast-2-Hybrid, Tandem Affinity Purification, Protein Microarrays, Co-immunoprecipitation experiments etc.

You can find a list of your uploaded biological networks. In the list, you can view the file name, the file description, the type and the format of the file, the file size, the date of the file creation and the date of the last modification. You can always download or delete each file listed.

For a Biological Network file type you can select one of the four available actions:

- Visualization at BioNets,
- Network Clustering at BioNets,
- Network Analysis at BioNets, or
- Networks Comparison at BioNets

and be guided to the respective functionality of BioNets.

For the Networks Comparison at BioNets action, you also have to select a second biological network file. The uploaded network could be weighted. That means that each edge of the network must have a weight assigned. If the network is unweighted, you should assign equal weight (for example 1) to all edges.

BIOLOGICAL NETWORK CLUSTERS FILES

You can upload a new Biological Network Clusters file. The Biological Network Clusters datasets which can be handled by the current version are those derived from experimental techniques for the prediction of protein-protein interactions and protein complexes such as Yeast-2-Hybrid, Tandem Affinity Purification, Protein Microarrays, Co-immunoprecipitation experiments etc.

You can find a list of your uploaded biological network clusters. In the list, you can view the file name, the file description, the type and the format of the file, the file size,

the date of the file creation and the date of the last modification. You can always download or delete each file listed.

For a Biological Network Clusters file type you can select:

- Visualization at BioNets

and be guided to the respective functionality of BioNets.

SOFT FILES

You can upload a new SOFT file. Simple Omnibus Format in Text (SOFT) is the format supported by Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo/) and it is the prevalent format for gene expression experiments. It is a simple line-based, plain text format, meaning that SOFT files may be readily generated from common spreadsheet and database applications. A single SOFT file can hold both data tables and accompanying descriptive information for multiple, concatenated Platforms, Samples, and/or Series records. Affymetrix CHP files and supplementary data files like Affymetrix CEL or GenePix GPR scan files can also be provided - just bundle them together with the SOFT file into a .zip, .rar, or .tar archive using a program like WinZip at the time of submission.

You can find a list of your uploaded SOFT files. In the list, you can view the file name, the file description, the type and the format of the file, the file size, the date of the file creation and the date of the last modification. You can always download or delete each file listed.

For a SOFT file you can select:

- Extract Gene Expression

and be guided to the respective functionality of BioNets.

MOLECULES QUANTIFICATION FILES

You can upload a new gene expression (or more generally a molecules quantification) file. Gene expression files are tab delimited files in the following format:

gene_name_1 expr1 … exprn

gene_name_2 expr1 … exprn

…

You can find a list of your uploaded or saved from InSyBio BioNets Tool gene expression files. In the list, you can view the file name, the file description, the type and the format of the file, the file size, the date of the file creation and the date of the last modification. You can always download or delete each file listed.

For a gene expression file you can select one of the two available actions:

- Create Gene Expression Network, or
- Differential Expression Biomarkers from Analyzing Gene Expressions

and be guided to the respective functionality of BioNets.

For the Differential Expression Biomarkers from Analyzing Gene Expressions action, you also have to select a second gene expression file.

DIFFERENTIAL EXPRESSION BIOMARKERS FILES

You can upload a new differential expression biomarkers file. Differential expression biomarkers files are tab delimited txt files with two columns: the first column includes the names of the biomarkers and the second their p-values.

You can find a list of your uploaded or saved from InSyBio BioNets Tool differential expression biomarkers files. In the list, you can view the file name, the file description,

the type and the format of the file, the file size, the date of the file creation and the date of the last modification. You can always download or delete each file listed.

For a differential expression biomarkers file you can select:

- Merge it with network comparison biomarkers

and be guided to the respective functionality of BioNets.

For the Merge biomarkers from network comparison biomarkers and differential expression biomarkers files action, you also have to select a network comparison biomarkers file.

NETWORK COMPARISON BIOMARKERS FILES

You can upload a new network comparison biomarkers file. Network comparison biomarkers files include the network comparison biomarkers. These biomarkers have been inferred by InSyBio's novel approach for comparing biological networks and locating nodes whose roles in the network has been significantly altered. Network comparison biomarkers files are tab delimited txt files with two columns: the first column includes the names of the biomarkers and the second a confidence score for this biomarker. This confidence score takes values from 0.5 to 1 and indicates the significance of the biomarker.

You can find a list of your uploaded or saved from InSyBio BioNets Tool network comparison biomarkers files. In the list, you can view the file name, the file description, the type and the format of the file, the file size, the date of the file creation and the date of the last modification. You can always download or delete each file listed.

For a network comparison biomarkers file you can select:

- Merge it with differential expression analysis biomarkers

and be guided to the respective functionality of BioNets.

For the Merge biomarkers from network comparison biomarkers and differential expression biomarkers files action, you also have to select a differential expression analysis biomarkers file.

MERGED BIOMARKERS FILES

You can upload a new merged biomarkers file. Merged biomarkers files include the biomarkers which have been inferred from merging differential expression biomarkers with network comparison biomarkers. Merged Biomarkers files are tab delimited txt files with two columns: the first column includes the names of the biomarkers and the second a confidence score for this biomarker. This confidence score is estimated using the p-value of the differential expression biomarkers and the confidence score of the network based biomarkers.

You can find a list of your uploaded or saved from InSyBio BioNets Tool merged biomarkers files. In the list, you can view the file name, the file description, the type and the format of the file, the file size, the date of the file creation and the date of the last modification. You can always download or delete each file listed.

DATASETS FROM INSYBIO INTERACT

You can find a list with your saved PPI datasets from InSyBio Interact Tool. In the list, you can view for each dataset, the number of positive and negative PPIs, if the features are normalized or not, if the positive examples of the dataset are in HPRD database and the date that the dataset was created. You can always download or delete your dataset.

From the InSyBio Tools Actions menu, you can be guided to Interact in order to view and edit your dataset.

BIOMARKERS DATASETS

You can upload a new biomarkers dataset. Biomarker datasets are tab delimited txt files. The rows are the biomarkers (molecules, or else features) and the columns are the samples. Thus, each cell contains the quantification value of a biomarker in a sample. Files with and without headers are both supported.

You can find a list of your uploaded or saved from InSyBio Biomarkers Tool biomarkers datasets. In the list, you can view the file name, the file description, the type and the format of the file, the file size, the date of the file creation and the date of the last modification. You can always download or delete each file listed.

BIOMARKERS LABELS FILES

You can upload a new biomarkers labels files. Biomarker labels are tab delimited txt files. The labels (each one corresponding to a sample) should be all in one row, separated by tabs.

You can find a list of your uploaded or saved from InSyBio Biomarkers Tool biomarkers labels files. In the list, you can view the file name, the file description, the type and the format of the file, the file size, the date of the file creation and the date of the last modification. You can always download or delete each file listed.

BIOMARKERS MODEL FILES

You can find a list with your saved model files from InSyBio Biomarkers Tool. You can always download or delete your dataset.

From the InSyBio Tools Actions menu, you can be guided to Biomarkers in order to apply the trained model to new data.

How to get InSyBio Datastore

To request a free one month full (evaluation) version of InSyBio Suite please email us at info@insybio.com .

To purchase InSyBio Datastore commercial version 2.4 please contact us at sales@insybio.com.

About Us

InSyBio Ltd is a bioinformatics pioneer company (www.insybio.com) in personalized healthcare, that focuses on developing computational frameworks and tools for the analysis of complex life-science and biological data in order to develop predictive integrated biomarkers (biomarkers of various categories) with increased prognostic and diagnostic aspects for the personalized Healthcare Industry.

InSyBio Suite consists of tools for providing integrated biological information from various sources, while at the same time it is empowered with robust, user-friendly and installation-free bioinformatics tools based on intelligent algorithms and methods.

COPYRIGHT NOTICE