**Aim**

*A good example is the best sermon*, claimed American polymath Benjamin Franklin. This worked-out example is intended to illustrate from scratch the steps that must be followed to attain the different statistical calculations that *iSanXoT* can provide.

**Project description and starting files**

|  |  |
| --- | --- |
| Table I. **Isotopic sample labeling.** | |
| **Sample Name** | **iTRAQ label** |
| Jurkat WT 1 | 113 |
| Jurkat WT 2 | 114 |
| Jurkat WT 3 | 115 |
| Jurkat WT 4 | 116 |
| Jurkat KO 1 | 117 |
| Jurkat KO 2 | 118 |
| Jurkat KO 3 | 119 |
| Jurkat KO 4 | 121 |

We shall work with a simple model consisting of four wild-type (WT) and four knock-out (KO) Jurkat cell preparations to be compared. The protein extracts obtained from these eight samples were subjected to tryptic protein digestion, after which the resulting peptides were isotopically labelled using the iTRAQ Reagent 8 plex kit[[1]](#footnote-1) as depicted in Table I. Then the labeled peptide samples were pooled and separated into five fractions using high-pH reversed-phase chromatography[[2]](#footnote-2). Liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) analysis of both the unfractionated material and the five peptide fractions resulted in the six raw LC-MS/MS files that can be found in …\WOE\Raw\_files. These raw files were used as inputs to Protein Discoverer 2.1[[3]](#footnote-3) to generate, among others, the six *PSM.txt* files stored in …\WOE\PSM.txt\_files upon database searching against the FASTA human protein database that can be found in …\WOE\Database\_and\_category\_files. Please note that this protein database contains both true protein sequences and the corresponding inverted (decoy) sequences to enable false discovery rate (FDR) estimation for peptide identification.

A close inspection of the *PSM.txt* files reveals that these plain text files hold 57 columns displaying information for every peptide-spectrum match (PSM)

|  |  |
| --- | --- |
| Table II. **Data accessed by iSanXoT in the *PSMs.txt* files.** | |
| **Header** | **Description** |
| Spectrum File | Name of the raw LC-MS/MS file |
| First Scan | Scan number identifier |
| Sequence | Amino acid peptide sequence |
| Modifications | Unimod[[4]](#footnote-4) chemical or posttranslational modifications to peptide sequence |
| Charge | Peptide charge |
| XCorr | Cross-correlation value as provided by the SEQUEST[[5]](#footnote-5) algorithm |
| MH+ [Da] | Measured monoisotopic protonated peptide mass in Da |
| Theo. MH+ [Da] | Theoretical monoisotopic protonated peptide mass in Da |
| DeltaM [ppm] | Difference between measured and theoretical monoisotopic mass in ppm |
| Protein Accessions | Accession codes for the proteins to which the peptide sequence is ascribed |
| 113-121 | Intensity of iTRAQ reporter ions 113-121 |

obtained upon database searching; however, only a reduced subset of these data will be accessed by *iSanXoT,* as shown in Table II.

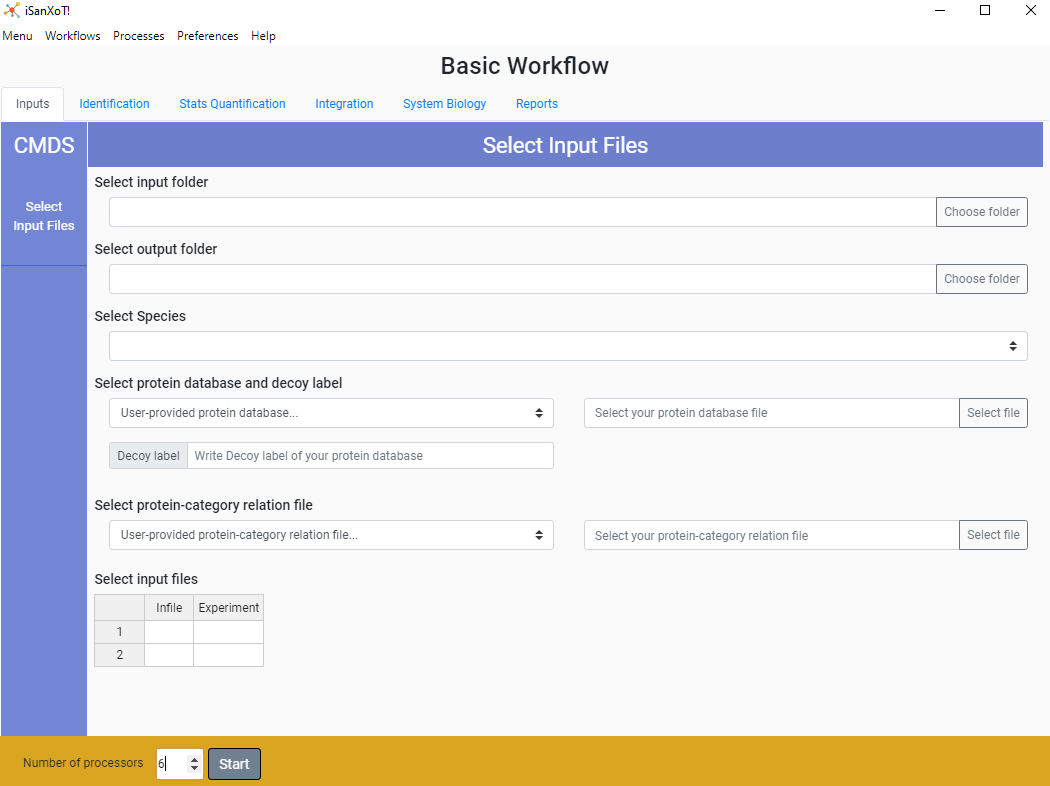
**Getting started: the *Inputs* menu**

Once *iSanXoT* has been successfully installed, executing the *isanxot.bat* script will bring you to *iSanXoT*’s main page. Select *Basic Workflow* here and then *Go to workflow with sample data*. You will be taken to the *Basic Workflow Inputs* window (Fig. 1), where the location of the input files necessary for *iSanXoT* to work with the sample data is specified:

* + *Select input folder* specifies the location of the files containing PSM identification and quantification data (the *PSMs.txt* files in this example);
  + *Select output folder* describes the path to the folder where *iSanXoT* output files will be stored. Selection of an *output folder* other than the *input folder* is strongly recommended;
  + *Select species* indicates the organism(s) from which the samples were derived (human in this case)*;*
  + *Select protein database and decoy label* allows the user to specify which FASTA protein database was used for peptide identification (the one stored in …\WOE\Database\_and\_category\_files in this case). Alternatively, one of the FASTA databases available through *iSanXoT* can be selected, but be aware that these will not exactly match the protein database you actually used. In the *Decoy label* field write the text tag that is used in the FASTA database to designate decoy (inverted) protein sequences (INV\_ for the sample data);

Figure 1. **The Basic Workflow Inputs window after loading the sample data**.

* + *Select protein-category relation file* enables selection of the protein-category relation file to be used by *iSanXoT* for the statistical calculations based on the Systems Biology Triangle[[6]](#footnote-6)



1. Ross P.L. *et al*. (2004) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents*. Mol. Cell. Proteomics* **3**, 1154–69. [↑](#footnote-ref-1)
2. Stein D.R. *et al*. (2013) High pH reversed-phase chromatography as a superior fractionation scheme compared to off-gel isoelectric focusing for complex proteome analysis. *Proteomics* **13**, 2956-66. [↑](#footnote-ref-2)
3. https://www.thermofisher.com/order/catalog/product/OPTON-30810#/OPTON-30810. [↑](#footnote-ref-3)
4. https://www.unimod.org/ [↑](#footnote-ref-4)
5. Eng J.K *et al.* (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom*. **5**, 976–89. [↑](#footnote-ref-5)
6. García-Marqués et al., [↑](#footnote-ref-6)