SUPPLEMENTARY MATERIAL

iSanXoT: a standalone application for the integrative analysis of mass spectrometry-based quantitative proteomics data

Jose Manuel Rodríguez[1], Inmaculada Jorge[2,3], Ana Martinez-Val[2], Rafael Barrero-Rodríguez[2], Ricardo Magni[2,3], Estefanía Núñez[2,3], Andrea Laguillo[2], Cristina A. Devesa[2], Juan A. López[1,3], Emilio Camafeita[1,3,*], Jesús Vázquez[2,3,*]

[1] *Proteomics Unit, Centro Nacional de Investigaciones Cardiovasculares Carlos III (CNIC), 28029 Madrid, Spain*

[2] *Cardiovascular Proteomics Laboratory, Centro Nacional de Investigaciones Cardiovasculares Carlos III (CNIC), 28029 Madrid, Spain*

[3] *CIBER de Enfermedades Cardiovasculares (CIBERCV), Madrid, 28029, Spain*

[*] To whom correspondence should be addressed.

Table of Contents

Sample Workflows with Application to Case Studies

Below, we provide detailed descriptions of four sample workflows that illustrate the capability of iSanXoT to statistically ascertain changes in protein or peptide abundance across various biological contexts. It is important to note that these workflows can be easily reused to process new data (refer to the next section).

Workflow 1: One-step quantification in a labeled experiment

Experimental

The identification and quantification data from García-Marqués *et al*. [1] were used to illustrate this workflow. This study characterizes the molecular alterations that take place over time when vascular smooth muscle cells (VSMCs) are treated with angiotensin-II (AngII) for 0, 2, 4, 6, 8, and 10 hours. Quantitative proteomics were performed using isobaric iTRAQ 8-plex labeling. Workflow 1 analyzes a) changes in protein abundance and b) alterations in functional categories produced by the coordinated behavior of proteins at each of the specified times, in relation to time 0. This is achieved in only one step using the compound module WSPP-SBT, which automatically performs all the required tasks.

Workflow execution

The workflow template and required input files for executing this workflow can be downloaded from https://raw.githubusercontent.com/CNIC-Proteomics/iSanXoT/master/docs/templates/WSPP-SBT.zip

Refer to the *Importing a Workflow Template* section below for detailed instructions.

Workflow operation

Workflow 1 requires the RELS CREATOR module, the WSPP-SBT compound module, and the REPORT basic module (*Figure S1*). The relation tables necessary for performing the integrations are created by the RELS CREATOR module (*Figure S1*A) from a table provided by the user. The WSPP-SBT module performs a sequence of consecutive integrations based on the WSPP statistical model [2] and the SBT algorithm [1] (*Figure S1*B). Finally, the REPORT module organizes the data into tables containing the required information.

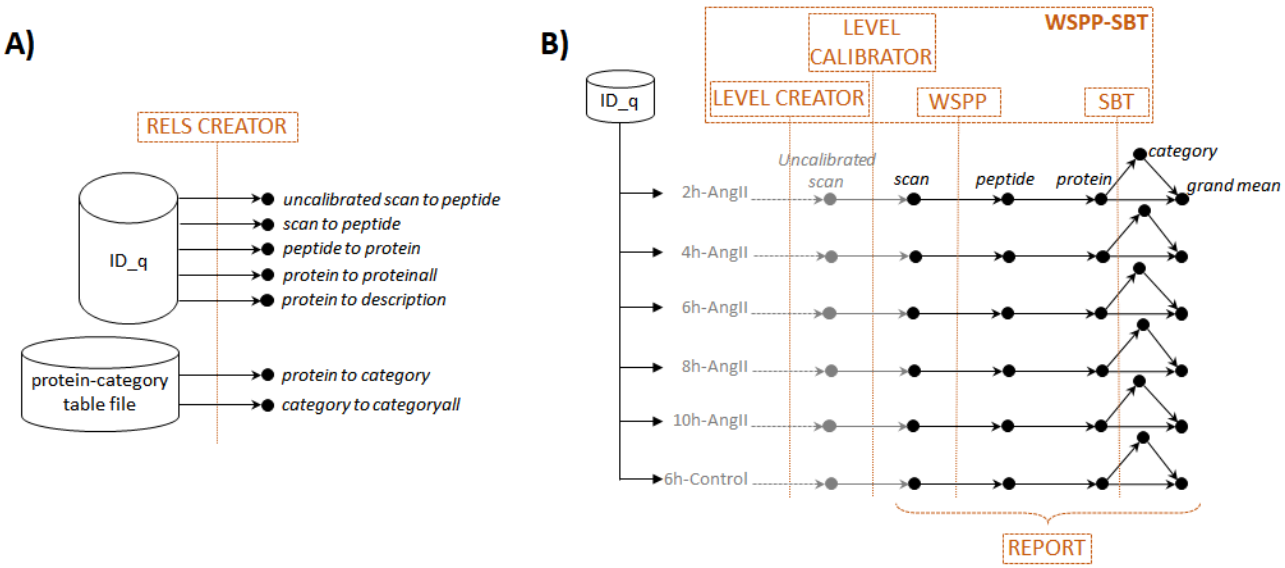


Figure S1. Scheme of workflow 1 (one-step quantification in a labeled experiment) showing module components: RELS CREATOR (A) and WSPP-SBT and REPORT (B)

The WSPP-SBT module requires the user to define the meaning of relative abundances, which iSanXoT consistently expresses as log2ratios. In this case, the abundance data corresponds to the intensities of iTRAQ reporters at the scan level, tabulated in the "ID-q" file with the name of each reporter as a column header (see below for how these tables are generated). The intensities of each scan at 0 h are in the "Abundance: 113" column and serve as a common reference to express abundance ratios; thus, they are used as the denominator. The reporter intensities corresponding to different time points serve as numerators for the ratios. The task table also enables the user to assign an easily identifiable name to the folders where the quantitative values of each sample are stored (*Figure S2*).

Experiment	Identifier column header	Ratio numerator	Ratio denominator	Output Sample folder
VSMC	Scan_Id	Abundance: 114	Abundance: 113	2h-AngII
VSMC	Scan_Id	Abundance: 115	Abundance: 113	4h-AngII
VSMC	Scan_Id	Abundance: 116	Abundance: 113	6h-AngII
VSMC	Scan_Id	Abundance: 117	Abundance: 113	6h-Control
VSMC	Scan_Id	Abundance: 118	Abundance: 113	8h-AngII
VSMC	Scan_Id	Abundance: 119	Abundance: 113	10h-AngII

Figure S2. The WSPP-SBT task table for workflow 1.

The WSPP-SBT module initially conducts a calibration process, assigning a statistical weight to each log2ratio value at the scan level (*Figure S1B*), as described [2]. The statistical weight of each scan is the inverse of the estimated variance associated with the log2 of intensity ratios [2]. Following the

calibration of data at the scan level, the workflow proceeds with integrations from *scan-to-peptide* and *peptide-to-protein*.

At the protein level, the SBT algorithm is applied for the detection of functional category changes originating from the coordinated behavior of proteins (*Figure S1*B). The algorithm first calculates the variance of the *protein-to-category* integration, providing an improved estimate of the technical protein variance that is less influenced by biological changes [1]. This protein variance is then utilized to perform the *protein-to-grand* mean integration (hereinafter referred to as *protein-to-proteinall*), from which statistically significant abundance changes are detected. Finally, the algorithm conducts the *category-to-grand* mean integration (hereinafter referred to as *category-to-categoryall*), identifying statistically significant category changes. All results from the integrations performed by the WSPP-SBT module are saved for each sample in the *Output Sample folder*, as indicated in the module task table (*Figure S2*).

In every integration step, a relation table (a text file) is required to link lower- to higher-level elements. These relation tables can be automatically generated by the RELS CREATOR module (*Figure S1*A, upper) or provided by the user (*Figure S1*A, lower). In this example (*Figure S3*) the relation tables linking scans to peptides and peptides to proteins are obtained from the "ID-q" file by specifying the column names where they are located (in this case, "Scan_Id", "Pep_Id", and "Master Protein Accessions"). The columns "Master Protein Accessions" and "Master Protein Descriptions" in the "ID-q" file contain the accession numbers and complete names of the proteins, respectively. Consequently, a relation table, *protein2description*, is also created, which can later be used to append the full name of the protein to any of the created reports (see below).

An example of the *peptide2protein* relation table, linking the identified peptides to the proteins they originate from, is shown *Figure S4*A. The elements of the relation table *protein2category* were extracted from a text file containing functional annotations for mouse proteins, compiled from various protein function databases (*Figure S4*B), as described by the authors [1]. It is important to note that relation tables are by default extracted from the "ID-q" file. To use other text files, the absolute path with the location of the text file must be indicated. The relation tables *protein2proteinall* and *category2categoryall* guide the integration to a grand mean (a common element called "[1]"). Although the integration *peptide2peptideall* is not necessary in this workflow, it is included in this example since it may be useful to inspect quantifications at the peptide level.

Relation Table to be created	Column name of Lower level	Column name of Higher level	Column name of 3rd c	Table from which RT is extracted
uscan2peptide	Scan_Id	Pep_Id		
scan2peptide	Scan_Id	Pep_Id		
peptide2protein	Pep_Id	Master Protein Accessions		
peptide2peptideall	Pep_Id	[1]		
protein2proteinall	Master Protein Accessions	[1]		
protein2description	Master Protein Accessions	Master Protein Descriptions		
protein2category	Protein	Category		{PATH}/DAVID_IPA_merged_noDups_feb14_IJ.txt
category2categoryall	Category	[1]		{PATH}/DAVID_IPA_merged_noDups_feb14_IJ.txt

Figure S3. The RELS CREATOR task table was specifically designed for workflow 1.

A)

protein	peptide
P23242	KVAAGHELQPLALVDQRPSSR__N-Term(iTRAQ8plex);K1(iTRAQ8plex)
P17182	AAVPSGASTGLYEALELRDNDKTR__N-Term(iTRAQ8plex);K22(iTRAQ8plex)
P23780	AGATLDLLVENMGR__N-Term(iTRAQ8plex)
A2AIM4; P58771-2; P58774-2; E9Q45(	RLQLVEEELDRAQER__N-Term(iTRAQ8plex)
P20152	EKLQEEMLQREEAESTLQSFR__N-Term(iTRAQ8plex);K2(iTRAQ8plex)
P99024	MAVTFLGNSTALQELFKR__N-Term(iTRAQ8plex);M1(Oxidation);K17(iTRAQ8plex)
A1BN54; Q9JI91; P57780; Q7TPR4	KHEAFESDLAAHQDR__N-Term(iTRAQ8plex);K1(iTRAQ8plex)
Q91YQ5	ASSFVLALEPELESR__N-Term(iTRAQ8plex)
O08547	NLGSLNTELQDVQR__N-Term(iTRAQ8plex)
Q9EQ06	SVAGELVLLTGAGHGLGR__N-Term(iTRAQ8plex)
P05064; Q9CPQ9	FSNEELAMATVTALR__N-Term(iTRAQ8plex)
P10126	KDGSASGTTLLEALDCLLPPTRPTDKPLR__N-Term(iTRAQ8plex);C16(Carbamidomethyl);K26(iTRAQ8plex)

B)

category	protein
DAVID_PANTHER_BP_ALL_BP00141:Transport	Q9Z351
DAVID_PANTHER_BP_ALL_BP00142:Ion transport	Q9Z351
DAVID_PANTHER_BP_ALL_BP00143:Cation transport	Q9Z351
DAVID_PANTHER_PATHWAY_P00042:Muscarinic acetylcholine receptor 1 and 3 s	Q9Z351
DAVID_SP_PIR_KEYWORDS_alternative splicing	Q9Z351
DAVID_SP_PIR_KEYWORDS_ion transport	Q9Z351
DAVID_SP_PIR_KEYWORDS_ionic channel	Q9Z351
DAVID_KEGG_PATHWAY_mmu04020:Calcium signaling pathway	Q9Z329
DAVID_KEGG_PATHWAY_mmu04070:Phosphatidylinositol signaling system	Q9Z329
DAVID_KEGG_PATHWAY_mmu04114:Oocyte meiosis	Q9Z329
DAVID_KEGG_PATHWAY_mmu04270:Vascular smooth muscle contraction	Q9Z329

Figure S4. An excerpt from the peptide2protein (A) and protein2category (B) relation tables, illustrating the links between peptides and proteins and proteins and categories, respectively.

After the integrations are executed, the REPORT module is employed to gather the specified statistical variables from the *Output Sample folders* designated by the user and organize them into tables (*Figure S5*). In this instance, the tabulation is focused on the results from the samples (2h-AngII, 4h-AngII, 6h-AngII, 8h-AngII, and 10h-AngII).

In this example the REPORT module creates a protein table and a category table by performing the following steps:

- Create a table named “Npep2prot”, which contains the count of peptides used to quantify each protein.
    - This involves extracting the number of elements (n) from the *peptide-to-protein* integrations in the specified folders, representing the lower level (peptide) used for the quantitation of the higher level (protein).

- Create a table named “Npep2prot_Quantprot_filtered”, including protein changes (Zqa) and their statistical significance (FDRqa).
    - This is achieved by extracting standardized log2 ratios (Z) and False Discovery Rates (FDR) from the *protein-to-proteinall* integration in the indicated folders, representing the lower level (protein).

- Add the count of peptides with which each protein is quantified to the “Npep2prot_Quantprot_filtered” table.
    - This step involves merging the previous table with the existing “Npep2prot” table based on the common level (protein), excluding a specific column (peptide), and eliminating

duplicate entries.

- Add an additional column to this table with the complete description of the proteins.
    - This is accomplished by merging the previous table with the relation table *protein2description* based on the common level in both tables (protein).

- Filter the table to include only proteins with a statistically significant abundance change (FDR < 0.01).
    - This is achieved by applying a condition based on the FDR to the results from the *protein2proteinall* integration. For more detailed information, refer to the "Filter for report" section in the iSanXoT documentation:
      https://cnic-proteomics.github.io/iSanXoT

- Create a table named "Nprot2cat", which contains the count of proteins used to quantify each category.
    - Extract the number of elements (n) from the *protein-to-category* integrations in the specified folders, representing the lower level (protein) used for the quantitation of the higher level (category).

- Create a table named "Nprot2cat_Quantcat_filtered", incorporating category changes (Zca) and their statistical significance (FDRca).
    - Extract standardized log2 ratios (Z) and False Discovery Rates (FDR) from the *category-to-categoryall* integration in the indicated folders, representing the lower level (category).

- Add to this table the count of proteins with which each category is quantified.
    - This is achieved by merging the previous table with the existing table "Nprot2cat" based on the common level in both tables (category), excluding a specific column (protein), and eliminating duplicate entries.

- Filter the table to include only categories with a statistically significant change (FDR < 0.01).
    - This is done by applying a condition based on the FDR to the results from the *category2categoryall* integration.

- Create a table named "Npep2prot_Quanprot" containing the number of peptides per protein, protein changes (Zqa), and their statistical significance (FDRqa).
    - This is done as previously explained, excluding the protein descriptions and filters.

- Create a table named "Nprot2cat_Quancat_Quanprot_filtered", including category changes (Zca) and their statistical significance (FDRca).
    - Extract standardized log2 ratios (Z) and False Discovery Rates (FDR) from the *category-to-categoryall* integration in the indicated folders, representing the lower level (category).

- Add to this table the count of proteins per category, protein changes (Zqa), and their statistical significance (FDRqa).
    - This is achieved by merging the previous table with the existing tables "Nprot2cat" and "Npep2prot_Quantprot".

- Filter the table to include only categories containing 5 or more proteins or 100 or fewer proteins.
    - This is done by applying a set of conditions in the *Filter* column, joined with the "&" operator.

Note that the commands in the REPORT module, facilitating the construction of tables essential for typical quantitative proteomics projects, are easily adaptable and reusable for other projects.

Sample folder(s)	Lower level	Higher level	Reported vars	Output report	Column headers to eliminate	Merge with report	Add columns from relation tal	Filter
2h-AngII , 4h-AngII , 6h-AngII , 8h-AngII , 10h-AngII	peptide	protein	n	Npep2prot				
2h-AngII , 4h-AngII , 6h-AngII , 8h-AngII , 10h-AngII	protein	proteinall	Z , FDR	Npep2prot_Quanprot_filtered	peptide	Npep2prot	protein2description	FDR_protein2proteinall < 0.01
2h-AngII , 4h-AngII , 6h-AngII , 8h-AngII , 10h-AngII	protein	category	n	Nprot2cat				
2h-AngII , 4h-AngII , 6h-AngII , 8h-AngII , 10h-AngII	category	categoryall	Z , FDR	Nprot2cat_Quancat_filtered	protein	Nprot2cat		FDR_category2categoryall < 0.01
2h-AngII , 4h-AngII , 6h-AngII , 8h-AngII , 10h-AngII	protein	proteinall	Z , FDR	Npep2prot_Quanprot	peptide	Npep2prot		
2h-AngII , 4h-AngII , 6h-AngII , 8h-AngII , 10h-AngII	category	categoryall	Z , FDR	Nprot2cat_Quancat_Quanprot_filtered		Nprot2cat , Npep2prot_Quanprot		(n_protein2category >=5) & (n_protein2category <= 100)

Figure S5. The REPORT task table that was designed for workflow 1.

In *Figure S7,* two heat maps are presented, constructed from the protein and category tables obtained using the REPORT module.

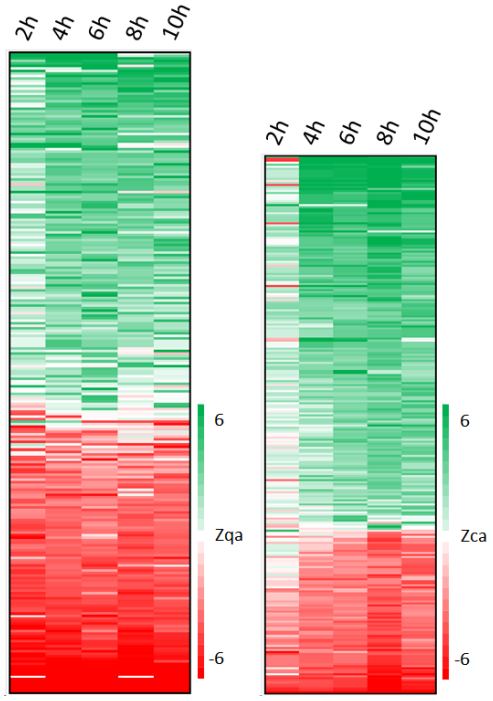


Figure S6. Relative abundance changes of proteins (Zqa, left) and functional categories (Zca, right) are derived from the "Npep2prot_Quanprot_filtered" and "Nprot2cat_Quancat_filtered" reports, respectively, generated by the REPORT module in workflow 1 (Figure S5). Both report tables were sorted based on the averages of Zqa and Zca, respectively.

In Figure S7 illustrates examples of functional categories displaying statistically significant changes resulting from coordinated protein behavior. The data for these plots are derived from the "Nprot2cat_Quancat_Quanprot_filtered" table generated by the REPORT module.

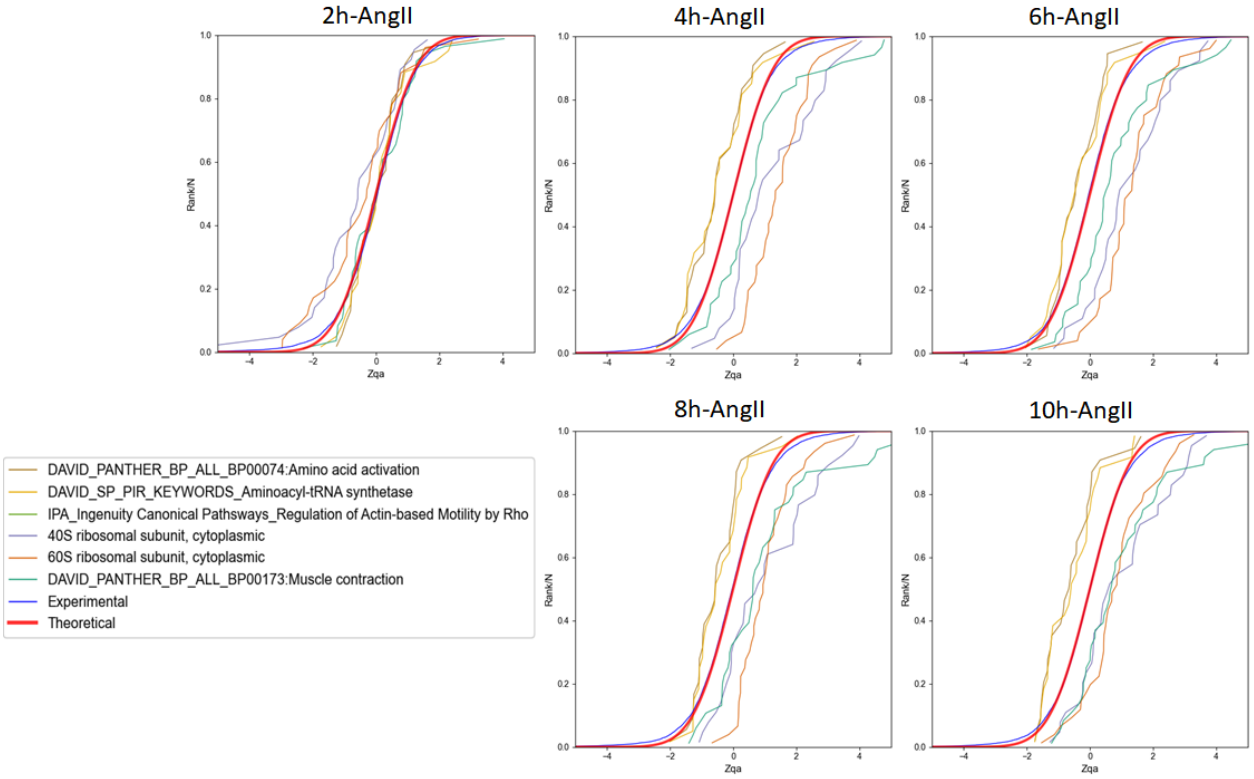


Figure S7. Examples of time-dependent coordinated protein behavior in VSMCs treated with angiotensin-II, revealed by the distribution of the standardized log2 ratio (Zqa) of protein components in each category.

Workflow 2: Step-by-step quantification and sample combination in a labeled experiment

Experimental

The workflow presented here utilizes data from González-Amor *et al*. [3] focusing on the contribution of interferon-stimulated gene 15 (ISG15) to vascular damage associated with hypertension. The study employs knockout mutants for the ISG15 gene, subjecting animals to AngII treatment or not. A total of 16 samples from mouse aortic tissue represent four groups: four WT-Control mice, four ISG15-KO mice, four WT+AngII mice, and four ISG15-KO+AngII mice. The experiment, conducted in two isobaric iTRAQ 8-plex batches, serves as a case study to guide the step-by-step creation of a workflow. This workflow integrates quantitative results from individual samples to the protein level, consolidates protein data across the four biological replicates in each group, establishes ratios between conditions, and analyzes functional category changes due to coordinated protein behavior using the SBT model.

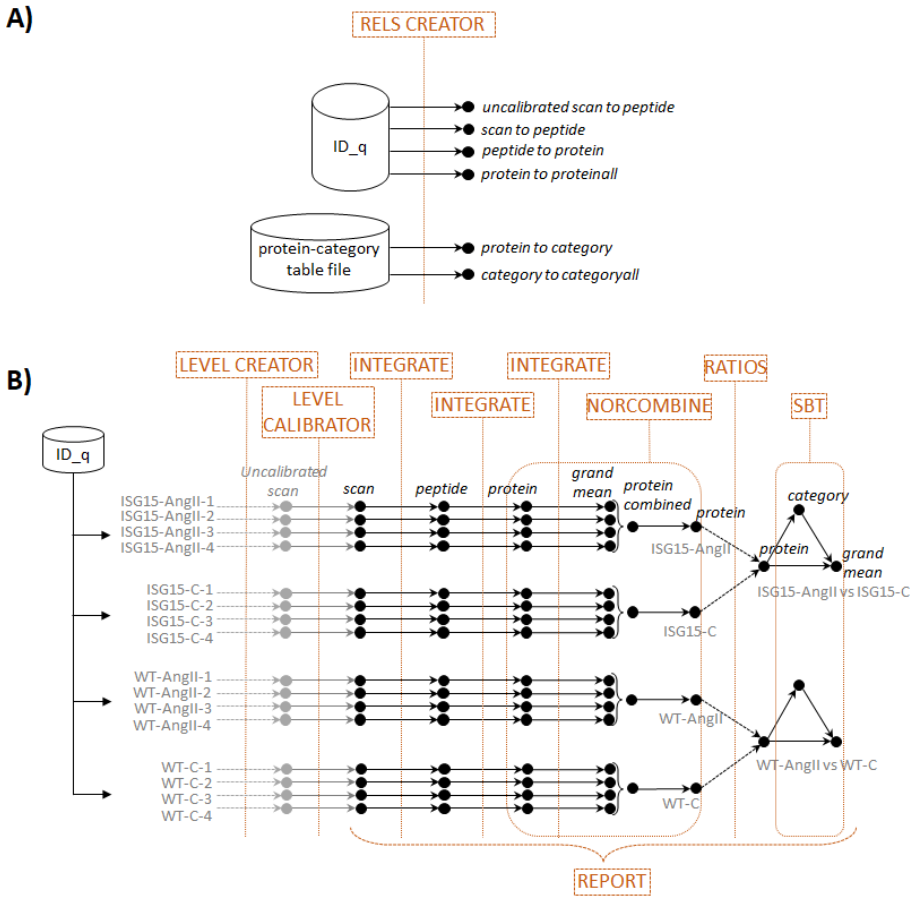


Figure S8. Schematic representation of workflow 2, which involves step-by-step quantification and sample combination in a labeled experiment. The figure illustrates the key modules in the workflow, including RELS CREATOR (A) and LEVEL CREATOR, LEVEL CALIBRATOR, INTEGRATE, NORCOMBINE, RATIOS, SBT, and REPORT (B)

Workflow execution

The workflow template and input files that are needed to execute this workflow can be downloaded from https://raw.githubusercontent.com/CNIC-Proteomics/iSanXoT/master/docs/templates/WSPP_NORCOM_RATIOS_SBT.zip

Refer to the *Importing a Workflow Template* section below for detailed instructions.

Workflow operation

Workflow 2 encompasses all six basic modules: LEVEL CREATOR, LEVEL CALIBRATOR, INTEGRATE, NORCOMBINE, RATIOS, and SBT, along with the REPORT module (*Figure S8*). The task table in the starting module, LEVEL CREATOR, generates files at the scan level containing log2 ratios and the corresponding sample folders (*Figure S9*). In this example, similar to workflow 1, the name of each iTRAQ reporter served as a column header in the "ID-q" file containing the intensities. Additionally, the "Experiment" column indicates whether the intensities come from the first or second iTRAQ 8-plex batch. Each iTRAQ batch comprises two biological replicates from each of the four groups. The average of reporter intensities from the two untreated WT mice (reporters in the columns "113" and "117") is used as an internal control within each batch, serving as the denominator for the log2 ratios.

Experiment	Identifier column header	Ratio numerator column	Ratio denominator column(s)	Level to be created	Output Sample folder
ISG15_iTRAQ1	Scan_Id	113	113, 117	u_scan	WT-C-1
ISG15_iTRAQ1	Scan_Id	114	113, 117	u_scan	WT-AngII-1
ISG15_iTRAQ1	Scan_Id	115	113, 117	u_scan	ISG15-C-1
ISG15_iTRAQ1	Scan_Id	116	113, 117	u_scan	ISG15-AngII-1
ISG15_iTRAQ1	Scan_Id	117	113, 117	u_scan	WT-C-2
ISG15_iTRAQ1	Scan_Id	118	113, 117	u_scan	WT-AngII-2
ISG15_iTRAQ1	Scan_Id	119	113, 117	u_scan	ISG15-C-2
ISG15_iTRAQ1	Scan_Id	121	113, 117	u_scan	ISG15-AngII-2
ISG15_iTRAQ2	Scan_Id	113	113, 117	u_scan	WT-C-3
ISG15_iTRAQ2	Scan_Id	114	113, 117	u_scan	WT-AngII-3
ISG15_iTRAQ2	Scan_Id	115	113, 117	u_scan	ISG15-C-3
ISG15_iTRAQ2	Scan_Id	116	113, 117	u_scan	ISG15-AngII-3
ISG15_iTRAQ2	Scan_Id	117	113, 117	u_scan	WT-C-4
ISG15_iTRAQ2	Scan_Id	118	113, 117	u_scan	WT-AngII-4
ISG15_iTRAQ2	Scan_Id	119	113, 117	u_scan	ISG15-C-4
ISG15_iTRAQ2	Scan_Id	121	113, 117	u_scan	ISG15-AngII-4

Figure S9. The LEVEL CREATOR task table specifically designed for workflow 2.

LEVEL CREATOR generates the *u_scan* (uncalibrated scan) files, comprising scan identifiers (extracted from the "Scan_Id" column in the "ID-q" table), log2-ratios at the scan level (*Xs*, as defined in the task table), and uncalibrated weights (*Vs*, corresponding to the intensities of the reporters in the Ratio

numerator column) (*Figure S10*). These uncalibrated weights (*Vs*) are indicative of quantification quality, where a higher weight implies more accurate quantification. However, at this stage, they are not associated with statistical variance.

Scan_Id	Xs_116_vs_113-117_Mean	Vs_116_vs_113-117_Mean
ISG15_iTRAQ1_v2-26378-2	0.166739599	354392.25
ISG15_iTRAQ1_v2-61753-2	0.513614912	574721.75
iTRAQ1_FR3-31890-3	0.055997478	51850.67578
ISG15_iTRAQ1_v2-62222-2	0.311700218	939171.75
iTRAQ1_FR2-28607-3	0.334866702	32351.34375
iTRAQ1_FR1_20190705161217-23524-2	0.516515769	2680378.5

Figure S10. An excerpt from one of the u_scan files generated by the LEVEL CREATOR module in workflow 2 is displayed, showing element identifiers in the left column, log2 ratios in the center column, and statistical weights in the right column.

The LEVEL CALIBRATOR module calibrates the Vs weights by conducting a *u_scan-to-peptide* integration, generating scan files with true, calibrated statistical weights, as defined in the WSPP model (the inverse of the estimated individual scan variances) (*Figure S11*, Top).

Sample f	Lower level for	Higher level for ir	Name of calibrated level
*	u_scan	peptide	scan

Sample folder(s)	Lower level	Higher level	Output Sample folder	Tag	FDR
*	scan	peptide			
*	peptide	protein			
*	protein	proteinall			0

Figure S11. The LEVEL CALIBRATOR (Top) and INTEGRATE (Bottom) task tables for workflow 2.

Note that the LEVEL CALIBRATOR automatically generates a plot to supervise the accuracy of calibrations (the "*_outGraph_VRank" PNG file) in each sample folder. This plot shows whether the model is able to predict experimental scan variances as a function of the calibrated statistical weights (see *Figure S12*).

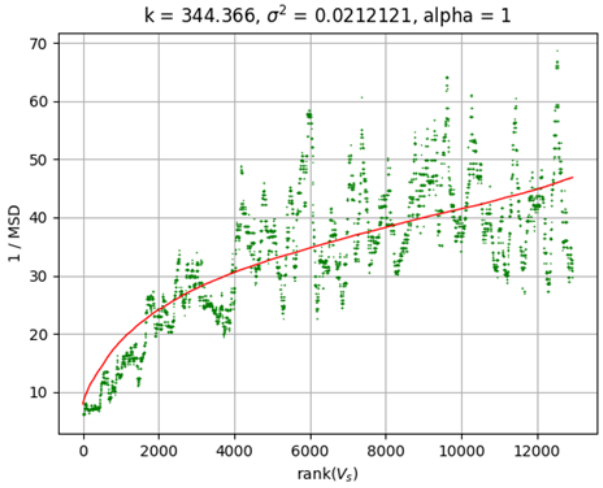


Figure S12. Automatically generated graphs to supervise the accuracy of calibrations. These graphs illustrate 1/MSD versus the rank of Vs (scan weight, which at this level corresponds to reporter intensity). MSD represents the Mean Squared Deviation of the scans versus the respective mean of the peptide to which they belong. The scans are ordered by Vs, and the MSD is calculated in a sliding window of 200 scans [2].

Similar to workflow 1, before conducting the integrations, the relation tables need to be created using the RELS CREATOR module. These tables exhibit a similar structure, as depicted in *Figure S13*.

Relation Table to be created	Column name of Lower level	Column name of Higher level	Column na	Table from which RT is extracted
u_scan2peptide	Scan_Id	Pep_Id		
scan2peptide	Scan_Id	Pep_Id		
peptide2protein	Pep_Id	Master Protein Accessions		
protein2proteinall	Master Protein Accessions	[1]		
protein2category	Protein	Category		{PATH}/uniprot-MusMusculus_GOanotationDavid_enero2018_rels_v8_Protein2Category.txt
category2categoryall	Category	[1]		{PATH}/uniprot-MusMusculus_GOanotationDavid_enero2018_rels_v8_Protein2Category.txt

Figure S13. Illustrates the RELS CREATOR task table for workflow 2.

The INTEGRATE module performs the *scan-to-peptide*, *peptide-to-protein*, and *protein-to-protein* integrations based on the module task table (*Figure S11*, Bottom). It's important to note that, for consistency, all files created for each sample are automatically stored in the folder specified in the task table of the LEVEL CREATOR module, unless an alternative location is specified in the *Output Sample folder* column.

Note that the INTEGRATE module automatically generates a plot to assess the accuracy of the GIA integration model in each integration step. This is achieved by comparing the distribution of Z values with that of the null hypothesis (standard normal distribution) (refer to *Figure S15*, left panels). These graphs are then stored in the respective sample folders as "*_outGraph.png" files.

By default, iSanXoT removes integration outliers. However, to avoid the removal of outlier elements in *the protein-to-proteinall* integration, particularly as these represent significantly altered proteins, a 0 FDR value was specified in the INTEGRATE task table for this integration (refer to Figure S11, Bottom).

Once protein levels are established, workflow 2 utilizes the NORCOMBINE basic module (*Figure S8B*) to integrate protein values from the four biological replicates within each group. The resulting integrated protein values per group are then stored in the folders WT-C, WT-AngII, ISG15-C, and ISG15-AngII (*Figure S14*).

Sample folders	Level	Norm	lowerNorm	Output Sample folder
WT-C-1 , WT-C-2 , WT-C-3 , WT-C-4	protein	proteinall	lowerNormV	WT-C
WT-AngII-1 , WT-AngII-2 , WT-AngII-3 , WT-AngII-4	protein	proteinall	lowerNormV	WT-AngII
ISG15-C-1 , ISG15-C-2 , ISG15-C-3 , ISG15-C-4	protein	proteinall	lowerNormV	ISG15-C
ISG15-AngII-1 , ISG15-AngII-2 , ISG15-AngII-3 , ISG15-AngII-4	protein	proteinall	lowerNormV	ISG15-AngII

Figure S14. The NORCOMBINE task table for workflow 2.

The NORCOMBINE module integrates biological replicates within sample groups using the GIA algorithm [1]. This algorithm models the distribution of protein values around the average, considering error propagation theory and estimating a global variance for the integration. The GIA algorithm operates under the assumption that individual variances of all lower elements (proteins) are influenced by a global variance, arising from biological variability within the same group. While this assumption may not hold in all the cases, it can be easily checked by inspecting the test distributions.

Similar to the INTEGRATE module, the NORCOMBINE module automatically generates graphs comparing the distribution of the integrated Z variables with those of the standard normal distribution. As depicted in *Figure S15* (right), the distribution of protein Z values estimated by the model in the case of the ISG15-AngII group aligns well with the null hypothesis. This agreement demonstrates that the assumption of the model is a suitable approach for handling the biological variance of samples within this group. Comparable results were obtained in the other three groups (not shown).

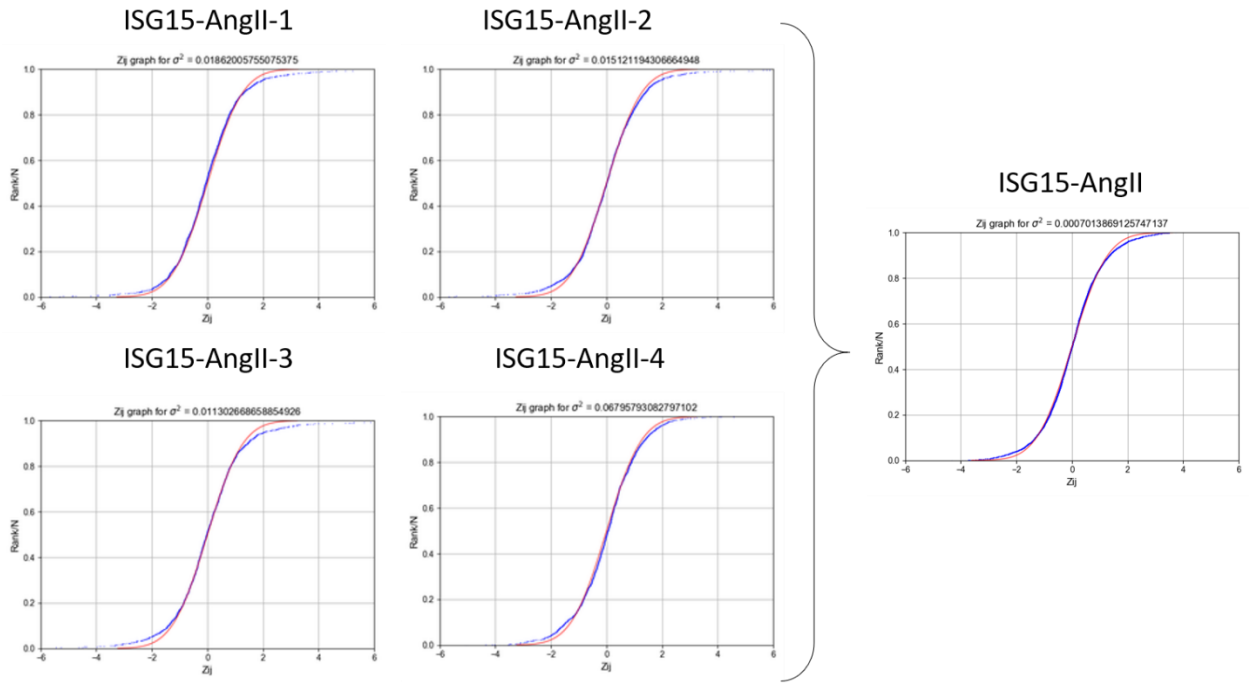


Figure S15. Distribution of the standardized log2 protein ratios (Zqa) from the four individual ISG15-AngII VSMC samples (Left panel). This illustrates how the WSPP model aligns well with the expected null distributions in the four cases. In the right panel, the figure shows the distribution from the integrated ISG15-AngII sample group obtained with the NORCOMBINE module. This demonstrates how the GIA assumption of a global biological variance is a valid approach to address the biological variability within this group. Red: null hypothesis (standard distribution); blue: experimental data.

Note also that the NORCOMBINE module employs a weighted averaging technique from multiple samples. Through its good fit to the null hypothesis (as illustrated in *Figure S15*), this approach enables accurate control over outliers. This unique approach allows the integration of protein values originating from unbalanced sample groups, distinct experiments, various mass spectrometers, and even different labeling techniques (refer to, for instance, [2]).

The module's task table specifies that samples were combined at the protein level using the *proteinall* level for normalization. In this process, log2 protein ratios are initially normalized by the grand mean before being integrated into an averaged protein value. This normalization compensates for differences in protein load into each iTRAQ channel. Importantly, it's worth noting that proteins could also be integrated at other levels (such as organelles, subcellular compartments, complexes, etc.) before being further integrated by NORCOMBINE, allowing for various types of normalization. Finally, the *lowerNorm* column indicates the file containing the normalized data, typically the *lowerNormV* files previously generated by the INTEGRATE module. For more detailed information, please refer to the iSanXoT documentation.

The protein averages derived from the four biological sample groups are then employed by the RATIOS basic module to calculate two ratios: *WT-AngIIvsWT-C*, where wild-type AngII-treated animals are compared to controls, and *ISG15-AngIIvsISG15-C*, where ISG15 AngII-treated animals are compared to ISG15 controls (*Figure S16*). In the "V method" column, users can specify the method used to assign a statistical weight to the log2ratios, with the default being the "max" method (for more details, please refer to the "RATIOS" module in the iSanXoT documentation: https://cnic-proteomics.github.io/iSanXoT).

Ratio numerator c	Ratio denominator	Level	V Method	Output Sample folder
WT-AngII	WT-C	protein	max	WT-AngIIvsWT-C
ISG15-AngII	ISG15-C	protein	max	ISG15-AngIIvsISG15-C

Figure S16. The RATIOS task table for workflow 2.

The final basic module executed in workflow 2 is the SBT (*Figure S17*), which applies the SBT algorithm to the previously defined comparisons. The goal is to detect changes in functional categories resulting from the coordinated behavior of proteins, following the approach explained in workflow 1. The SBT module offers greater flexibility by allowing triangle operations on any level, not just proteins. In this case, the triangle is formed by the levels protein and category (*Figure S16*) and the corresponding grand mean.

Sample folder(s)	Lower level	Intermediate level
WT-AngIIvsWT-C	protein	category
ISG15-AngIIvsISG15-C	protein	category

Figure S17. The SBT task table for workflow 2.

Finally, the REPORT module (*Figure S18*) is used, as in Workflow 1, to generate tables with protein and category data. In this case, additional features of the REPORT module are utilized. The table "Npep2prot" is generated using an asterisk. This symbol serves as a wildcard character for iSanXoT, indicating that the results from all samples containing peptide-to-protein integration (i.e., ISG15-AngII-1, ISG15-AngII-2, ISG15-AngII-3, ISG15-AngII-4, ISG15-AngII, ISG15-C-1, ISG15-C-2, ISG15-C-3, ISG15-C-4, ISG15-C, and ISG15-AngIIvsISG15-C) are to be included in the table.

However, the "Npep2prot_Quanprot_ISG15_filtered" and "Npep2prot_Quanprot_WT_filtered" tables include protein changes (Zqa), the statistical significance (FDRqa) of these changes, and the number of peptides per protein only from the samples indicated in the *Sample folder(s)* column. The report for the ISG15 samples is filtered by Zqa to display the most extreme values (greater than 1 or less than -1) but only for the "ISG15-AngIIvsISG15-C" sample. Additional filters for the minimum number of peptides per protein are also applied to these tables. The tables containing category values are filtered by Zca (greater than or equal to 2 or less than or equal to -2) and/or by the number of proteins per category (between 5 and 100).

Sample folder(s)	Lower level	Higher level	Reported vars	Output report	Column headers to elimi	Merge with report	Ad	Filter
*	peptide	protein	n	Npep2prot				
ISG15-AngII-1 , ISG15-AngII-2 , ISG15-AngII-3 , ISG15-AngII-4 , ISG15-AngII , ISG15-C-1 , ISG15-C-2 , ISG15-C-3 , ISG15-C-4 , ISG15-C , ISG15-AngIIvsISG15-C	protein	proteinall	Z , FDR	Npep2prot_Quanprot_ISG15_filtered	peptide	Npep2prot		(Z_protein2proteinall@ISG15-AngIIvsISG15-C <= -1 | Z_protein2proteinall@ISG15-AngIIvsISG15-C >= 1) & (n_peptide2protein >= 2)
WT-AngII-1 , WT-AngII-2 , WT-AngII-3 , WT-AngII-4 , WT-AngII , WT-C-1 , WT-C-2 , WT-C-3 , WT-C-4 , WT-C , WT-AngIIvsWT-C	protein	proteinall	Z , FDR	Npep2prot_Quanprot_WT_filtered	peptide	Npep2prot		(n_peptide2protein > 4)
ISG15-AngIIvsISG15-C , WT-AngIIvsWT-C	protein	category	n	Nprot2cat				
ISG15-AngIIvsISG15-C , WT-AngIIvsWT-C	category	categoryall	Z , FDR	Nprot2cat_Quancat_filtered	protein	Nprot2cat		(Z_category2categoryall <= -2 | Z_category2categoryall >= 2) & (n_protein2category >= 5) & (n_protein2category <= 100)
ISG15-AngIIvsISG15-C , WT-AngIIvsWT-C	protein	proteinall	Z , FDR	Nprot2cat_Quanprot		Nprot2cat		
ISG15-AngIIvsISG15-C , WT-AngIIvsWT-C	category	categoryall	Z , FDR	Nprot2cat_Quancat_Quanprot_filtered		Nprot2cat_Quanprot		(n_protein2category >= 5) & (n_protein2category <= 100)

Figure S18. The REPORT task table designed for workflow 2.

The tables generated by REPORT can be used to generate heatmaps showing the most relevant protein abundance changes (*Figure S19*). As previously shown [3], iSanXoT analysis revealed a coordinated alteration of proteins implicated in cardiovascular function, extracellular matrix and remodeling, and vascular redox state in aortic tissue from AngII-infused ISG15-KO mice (*Figure S20A*). The coordinated protein behavior from some of the altered categories can be analyzed in the sigmoid plots (*Figure S20B*).

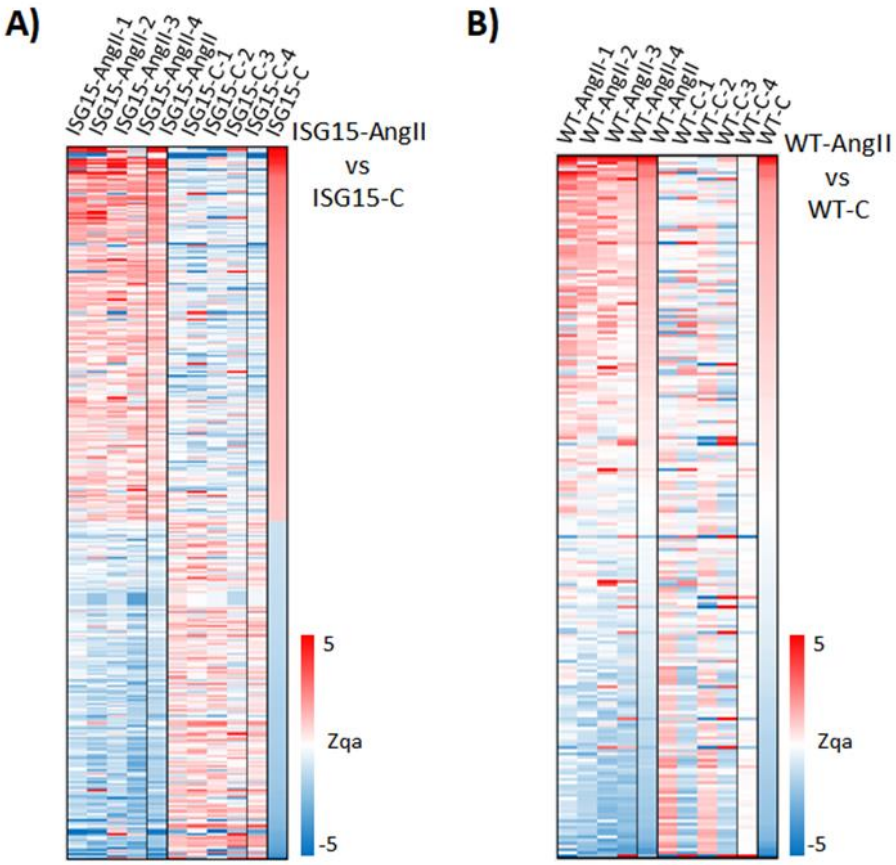


Figure S19. Differential abundance of functional proteins revealed by workflow 2. The heatmap (A) for proteins (Zqa) is based on the "Npep2prot_Quanprot_ISG15_filtered" REPORT table. The heatmap (B)

displays the proteins (Zqa) for the WT samples using the "Npep2prot_Quanprot_WT_filtered" REPORT table.

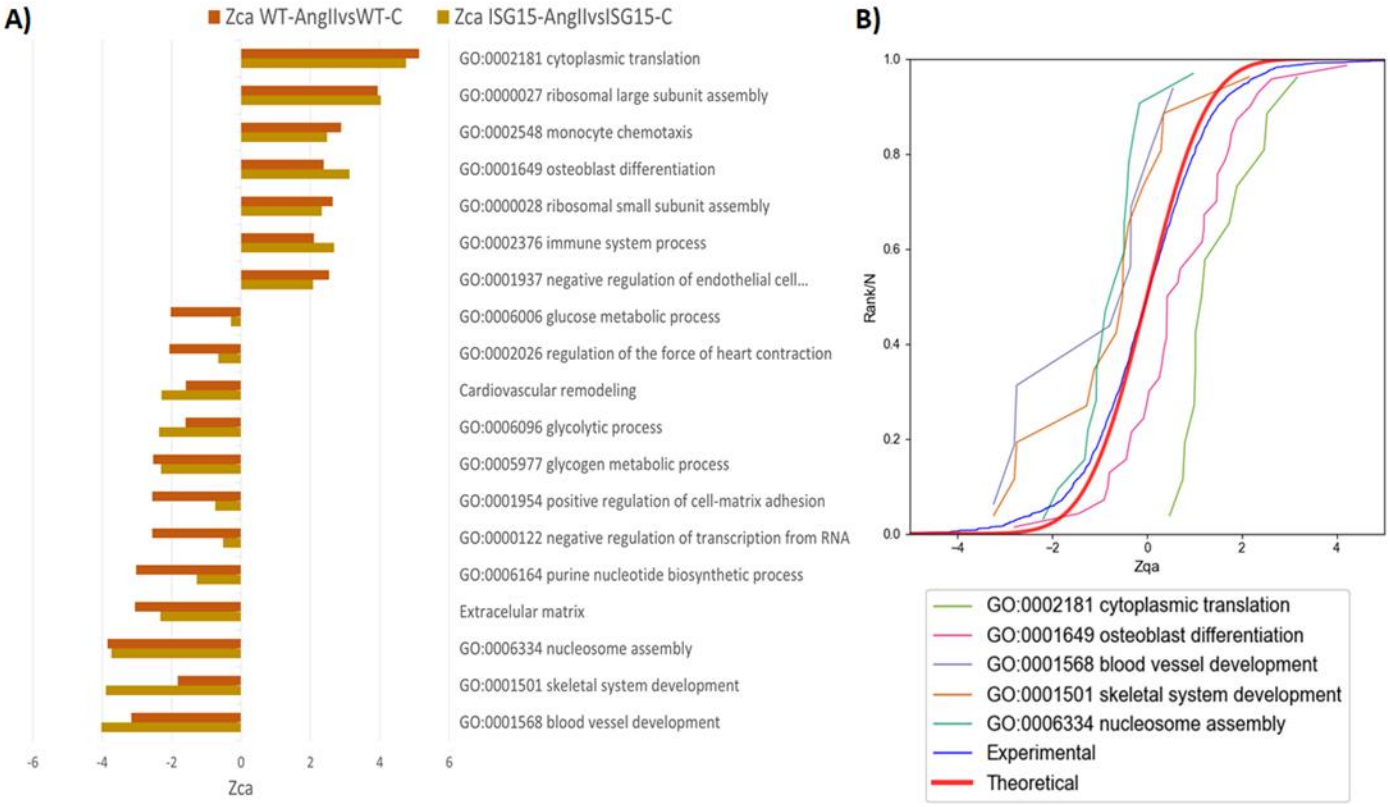


Figure S20. Functional category changes arising from coordinated protein behavior. A) Bar graph for functional categories (Zca) constructed from the "Nprot2cat_Quancat_filtered" REPORT table. B) The distributions of the standardized log2 protein ratios (Zqa) are shown for some of the functional categories that are significantly down-regulated (Left) or up-regulated (Right). The data to create the sigmoid curves are taken from the "Nprot2cat_Quancat_Quanprot_filtered" REPORT table.

Workflow 3: Quantification of posttranslationally modified peptides in a labeled experiment

Experimental

This workflow was employed to quantify reversibly oxidized Cys peptides in mouse embryonic fibroblast (MEF) preparations subjected to chemical oxidation with diamide. The experiment aimed to illustrate the comparative performance of on-filter (FASILOX) and in-gel (GELSILOX) approaches for studying the thiol redox proteome [4]. These techniques involved differentially labeling Cys residues based on their oxidation state, resulting in two distinct populations of reduced and oxidized Cys-containing peptides. MEF samples were incubated with diamide (treated group) or PBS (control group), and the resulting peptides were isobarically labeled with iTRAQ 8-plex (four biological replicates per condition). The workflow is designed to detect statistically significant abundance changes in peptides containing modified Cys residues.

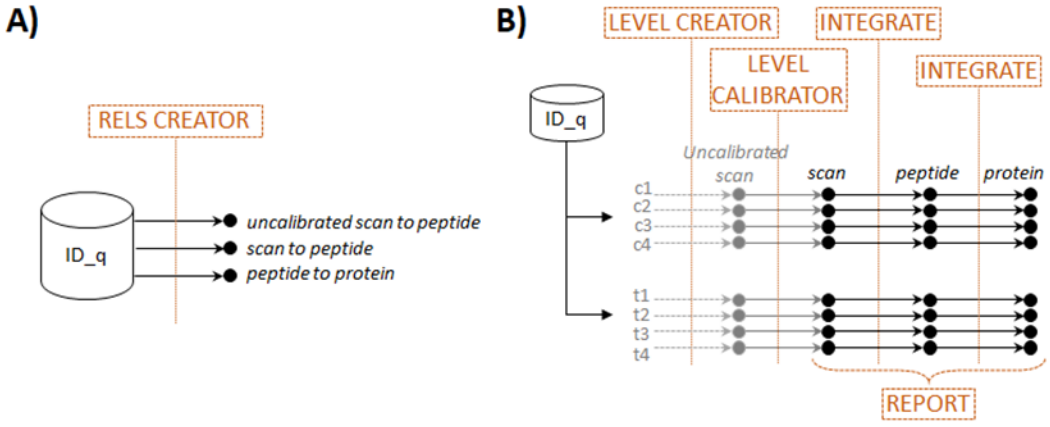


Figure S21. Scheme of workflow 3 (quantification of posttranslationally modified peptides in a labeled experiment) showing module components: RELS CREATOR (A) and LEVEL CREATOR, LEVEL CALIBRATOR, INTEGRATE, and REPORT (B).

Workflow execution

The workflow template and the required input files for executing this workflow can be downloaded from

https://raw.githubusercontent.com/CNIC-Proteomics/iSanXoT/master/docs/templates/WSPP_PTM.zip

Please refer to the *Importing a Workflow Template* section below for detailed instructions.

Workflow operation

Workflow 3 comprises the basic modules LEVEL CREATOR, LEVEL CALIBRATOR, and INTEGRATE, as well as the RELS CREATOR and REPORT modules (*Figure S21*) and is very similar to workflow 2. LEVEL CREATOR

was used to design the ratios and to generate the level files, sample folders and log2 ratios indicated in the corresponding task table (*Figure S22* and *Figure S23*). LEVEL CALIBRATOR was used to calibrate statistical weights (*Figure S24*, top) and INTEGRATOR to integrate from scan to peptide and from peptide to protein (*Figure S24*, bottom).

Experiment	Identifier column header	Ratio numerator column	Ratio denominator column(s)	Level to be created	Output Sample folder
JAL_FASILOX_iTR_ALL	ScanIdentifier	q_reporterion_113	q_reporterion_117 , q_reporterion_118 , q_reporterion_119 , q_reporterion_121	uscan	c1
JAL_FASILOX_iTR_ALL	ScanIdentifier	q_reporterion_114	q_reporterion_117 , q_reporterion_118 , q_reporterion_119 , q_reporterion_121	uscan	c2
JAL_FASILOX_iTR_ALL	ScanIdentifier	q_reporterion_115	q_reporterion_117 , q_reporterion_118 , q_reporterion_119 , q_reporterion_121	uscan	c3
JAL_FASILOX_iTR_ALL	ScanIdentifier	q_reporterion_116	q_reporterion_117 , q_reporterion_118 , q_reporterion_119 , q_reporterion_121	uscan	c4
JAL_FASILOX_iTR_ALL	ScanIdentifier	q_reporterion_117	q_reporterion_113 , q_reporterion_114 , q_reporterion_115 , q_reporterion_116	uscan	t1
JAL_FASILOX_iTR_ALL	ScanIdentifier	q_reporterion_118	q_reporterion_113 , q_reporterion_114 , q_reporterion_115 , q_reporterion_116	uscan	t2
JAL_FASILOX_iTR_ALL	ScanIdentifier	q_reporterion_119	q_reporterion_113 , q_reporterion_114 , q_reporterion_115 , q_reporterion_116	uscan	t3
JAL_FASILOX_iTR_ALL	ScanIdentifier	q_reporterion_121	q_reporterion_113 , q_reporterion_114 , q_reporterion_115 , q_reporterion_116	uscan	t4

Figure S22. The LEVEL CREATOR task table for workflow 3.

ScanIdentifier	Xs_q_reporterion_113_vs_q_reporterion_1	Vs_q_reporterion_113_vs_q_reporterion_1
JAL_FASILOX_iTR_ALL.raw-14205-3	-0.760406365	2154536.848
JAL_FASILOX_iTR_ALL.raw-19883-2	-0.64797195	475243.9143
JAL_FASILOX_iTR_ALL.raw-51554-3	-0.567309329	630711.4777
JAL_FASILOX_iTR_ALL.raw-77608-4	-0.620612557	826786.5206
JAL_FASILOX_iTR_ALL.raw-13670-2	-0.42826962	445258.3775
JAL_FASILOX_iTR_ALL.raw-50717-2	-0.490418129	324232.4633

Figure S23. Excerpt from one of the uscan *files generated by workflow 3 LEVEL CREATOR module showing element identifiers (left column), log2 ratios (center column) and statistical weights (right column).*

The only difference with workflow 2 lies in the INTEGRATE command used for the integration *peptide-to-protein*. INTEGRATE can use a modified version of the GIA algorithm for the quantitative analysis of posttranslational modifications (PTM) that includes a third column containing *tags* in the relation tables, as described [5]. In this workflow the advanced option of INTEGRATE was activated to display the *Tag* column, which is used to include only the peptides which are tagged in the relation table with the text "*Not modified"* when calculating the protein averages (*Figure S24*). An example of tagged *peptide2protein* relation table is shown in *Figure S25.* Proteins are thus quantified using only peptides which are not modified in Cys. However, although these Cys peptides do not contribute to protein averages, they are assigned a *Zpq* value, which serves to evaluate whether they deviate significantly from the expected distribution of peptides around their protein averages [4]. If the deviation is statistically significant it can be concluded that there is a change in abundance of the posttranslational modification in relation to the protein it comes from. This philosophy can be extended to any other kind of PTM.

Sample folder(s)	Lower level for integration	Higher level for integration	Name of calibrated level
*	uscan	peptide	scan

Sample folder(s)	Lower level	Higher level	Output Sample folder	Tag
*	scan	peptide		
*	peptide	protein		Not modified

Figure S24. The LEVEL CALIBRATOR (Top) and INTEGRATE (Bottom) task tables for workflow 3.

protein	peptide	Modifications
>tr|E9Q7Q3|E9Q7Q3_MOUSE Tropom	EQAEAEVASLNRR	Not modified
>tr|E9Q7Q3|E9Q7Q3_MOUSE Tropom	EQAEAEVASLNR	Not modified
>sp|Q8BTM8|FLNA_MOUSE Filamin-	SNFTVDC@SK{	Reduced-Cys peptides
>sp|P48962|ADT1_MOUSE ADP/ATP	DFLAGGIAAAVSK{	Not modified
>sp|Q8VDN2|AT1A1_MOUSE Sodium/	NLEAVETLGSTSTIC@SDK{	Reduced-Cys peptides
>sp|P20152|VIME_MOUSE Vimentin	QVQSLTC#EVDALK{	Oxidized-Cys peptides
>sp|Q9CPY7|AMPL_MOUSE Cytosol	QVIDC@QLADVNNLGK{	Reduced-Cys peptides
>sp|Q501J6|DDX17_MOUSE Probabl	GVEIC@IATPGR	Reduced-Cys peptides
>sp|Q9CZ44|NSF1C_MOUSE NSFL1 c	LGSTAPQVLNTSSPAQQAENEAK{	Not modified
>sp|B2RSH2|GNAI1_MOUSE Guanine	TTGIVETHFTFK{	Not modified

Figure S25. Excerpt from the peptide2protein relation table used to integrate peptides to proteins. Note the presence of a third column used to tag Cys-containing peptides, which will be excluded from the calculation of protein averages in the peptide-to-protein integration.

iSanXoT allows the automatic generation of relation tables containing tags, which are extracted from the "ID-q" table. To achieve this, RELS CREATOR employs a specific option (*Figure S26*). In this case, this option instructs RELS CREATOR to search the "ID-q" table for the column with the header *Modifications* and to translate its content into the third column of the *peptide2protein* relation table. In the "ID-q" table used in this instance, peptides containing modified Cys residues were labeled as either "Reduced-Cys peptides" or "Oxidized-Cys peptides" depending on the type of modification. RELS CREATOR locates these tags in the relation table (*Figure S25* and *Figure S26*). iSanXoT allows the use of any tag created by search engines or defined by the user, with the sole condition that the tag indicated in the INTEGRATE command must match the tag in the third column of the relation table (*Figure S24*, bottom).

Relation Table to be created	Column name of Lower level	Column name of Higher level	Column name of 3rd column
uscan2peptide	ScanIdentifier	Sequence	
scan2peptide	ScanIdentifier	Sequence	
peptide2protein	Sequence	FASTAshort	Modifications

Figure S26. The RELS CREATOR task table for workflow 3.

Finally, the REPORT module compiles the statistical variables generated by the *peptide-to-protein* integration for all the samples (*c1, c2, c3, c4, t1, t2, t3*, and *t4*), as indicated by the asterisk (*Figure S27A*). The REPORT commands mirror those employed in generating the protein tables in workflow 1, with the distinction that peptide values are tabulated instead of protein values, along with the inclusion of the number of scans per peptide instead of the number of peptides per protein. The false discovery rate (FDR) at the peptide level enables the detection of statistically significant changes in posttranslational modifications (PTM). This REPORT also generates a second filtered peptide table containing peptides with reduced Cys and the most pronounced abundance changes. This table was utilized to create a heatmap (*Figure S27B*).

A)

Sample fo	Lower level	Higher level	Reported vars	Output report	Column headers to elim	Merge with report	Ad	Filter
*	scan	peptide	n	Nscan2pep				
*	peptide	protein	Z , FDR, tags	Nscan2pep_Quanpepprot	scan	Nscan2pep		
*	peptide	protein	Z , tags	Nscan2pep_Quanpepprot_filtered	scan	Nscan2pep		(tags_peptide2protein == "Reduced-Cys peptides") & (Z_peptide2protein >= 2.5 | Z_peptide2protein <= -2.5)

B)

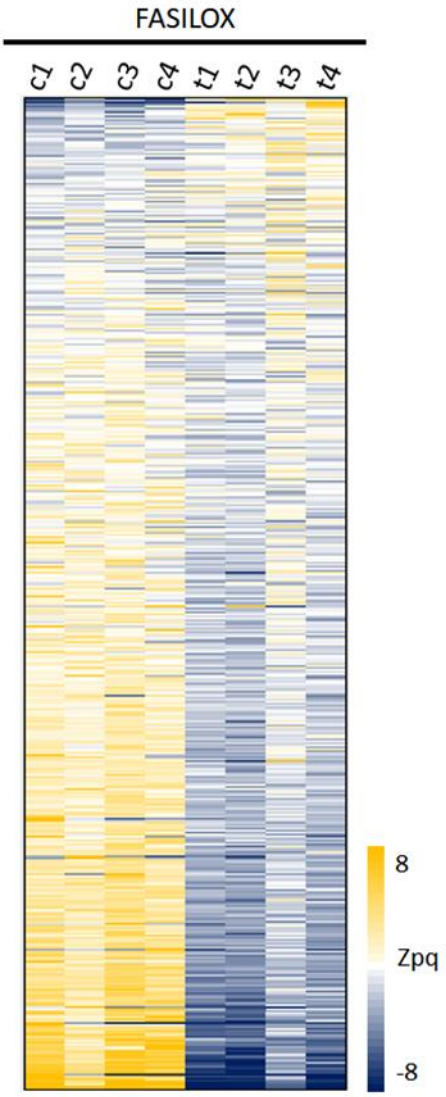


Figure S27. (A) The task table for Workflow 3 is in the REPORT module. (B) Relative abundance of Cys-containing peptides in MEF samples is represented by peptide log2 ratios expressed in units of standard deviation corrected by the protein mean (Zpq). The data for the heatmap were derived from the "Nscan2pep_Quanpepprot_filtered" report table.

Of significance, the peptides integrated followed a standard distribution in all eight samples, as depicted in blue in *Figure S28* for *t1*, *t2*, *c1*, and *c2* samples. This evidences that the error distribution at

the peptide level could be accurately modeled using the GIA algorithm. In addition, the treatment produced a generalized increase in the abundance of oxidized Cys-containing peptides (orange curves), with concomitant decrease in the abundance of reduced Cys-containing peptides (green curves). Consistently, the opposite changes were observed in the controls.

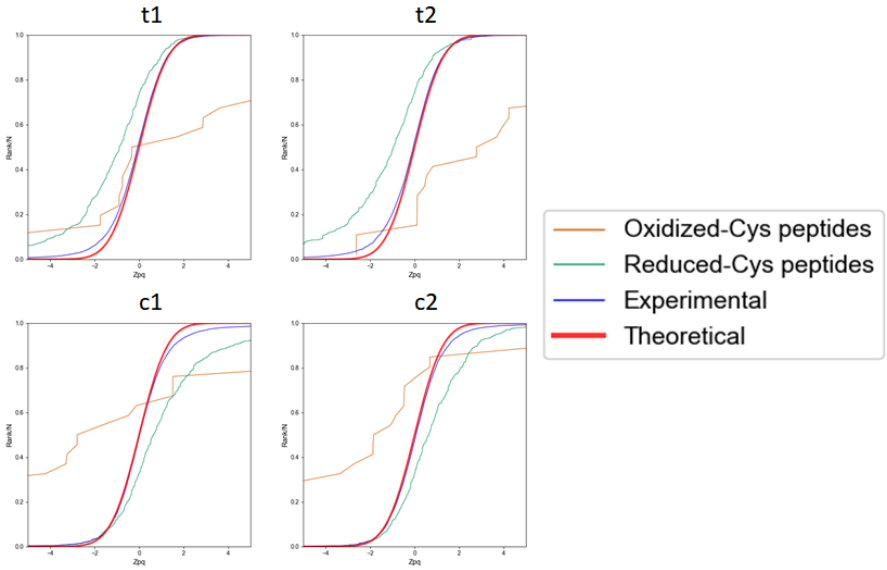


Figure S28. Distribution of the standardized variable at the peptide level (Zpq) in control MEF samples (t1, t2, c1 and c2) for all the peptides quantitated (blue) and the oxidized (orange) and reduced (green) Cys-containing peptide subpopulations. The theoretical normal distribution N(0,1) is shown in red. Positive/negative Zpq values indicate increased/decreased peptide abundance with respect to the average. These sigmoidal curves were created from the "Nscan2pep_Quanpepprot" table generated by REPORT.

Workflow 4: Label-free quantification

Experimental

This workflow was employed to analyze quantitative data obtained from a multicenter study conducted in Data-dependent Acquisition (DDA) mode. The study utilized samples prepared exactly as described in the paper by Navarro *et al*. [6]. Two hybrid proteome test samples were generated in the study, comprising tryptic digests of human, yeast, and *Escherichia coli* proteins mixed in two distinct proportions, as detailed in *Table S1*.

In this workflow, quadruplicate peptide preparations from each sample underwent analysis by LC-MS/MS. Following that, the MaxQuant [7] software was employed for peptide identification and quantification. For guidance on pre-processing data from MaxQuant and other software to adapt for use with iSanXoT, please consult the section described below.

Table S1. Proteome-hybrid samples A and B were prepared, each containing known quantities of peptide digestions of HeLa, Saccharomyces cerevisiae, and Escherichia coli. The samples were then mixed according to the procedure outlined in Navarro et al., Nature Biotech 2016.

	A	B	FOLD B/A
E. coli	20%	5%	0.25
S. Cerevisiae	15%	30%	2
HeLa	65%	65%	1

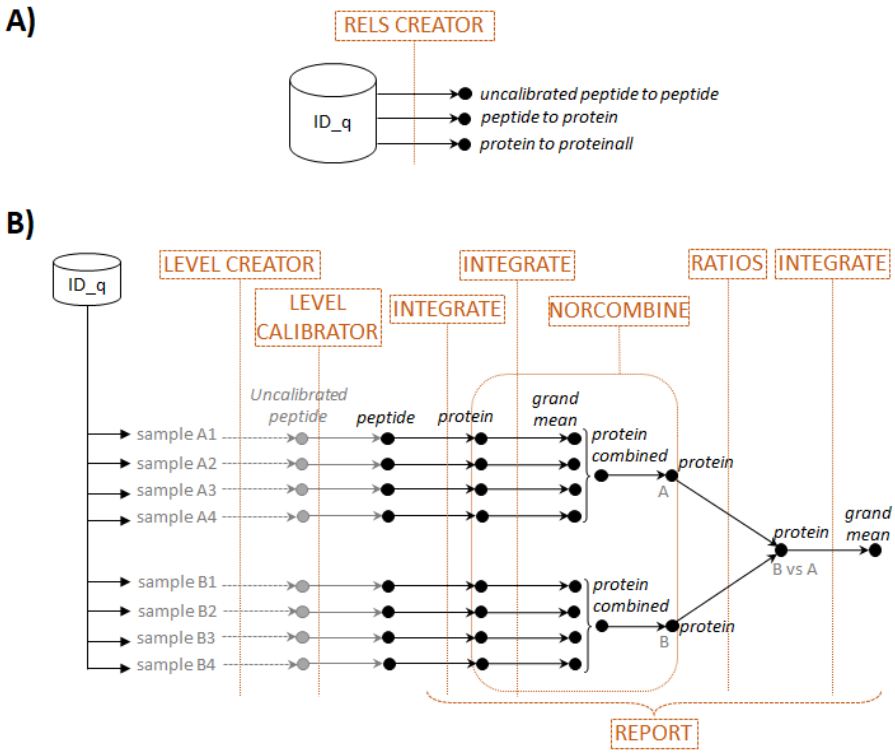


Figure S29. Scheme of workflow 4 (label-free quantification) showing module components: RELS CREATOR (A) and LEVEL CREATOR, LEVEL CALIBRATOR, INTEGRATE, NORCOMBINE, and REPORT (B).

Workflow execution

The workflow template and necessary input files for executing this workflow can be downloaded from

https://raw.githubusercontent.com/CNIC-Proteomics/iSanXoT/master/docs/templates/WPP_LabelFree.zip

Detailed instructions can be found in the *Importing a Workflow Template* section below.

Workflow operation

Workflow 4 includes the basic iSanXoT modules: LEVEL CREATOR, LEVEL CALIBRATOR, INTEGRATE, NORCOMBINE, and RATIOS, along with the REPORT and RELS CREATOR modules (*Figure S29*). The starting module, LEVEL CREATOR, generates the level files, sample folders and log2 ratios indicated in the corresponding task (*Figure S30*) based on the quantitative data at the peptide level obtained with MaxQuant for replicate A- and B-type samples. In this example, we used the average of peptide intensities

across all samples as the denominator of the log2 ratio. However, in this case, the averages of the four A-type and the four B-type samples are first calculated separately (as indicated by the square brackets), and then the average of the two averaged values is calculated (as indicated by the comma). This ensures that no log2 ratio is calculated when the four values are missing in either the A or the B sample group. This module generates uncalibrated files at the peptide level (*u_peptide*) (*Figure S31*).

Experiment	Identifier column header	Ratio numerator column	Ratio denominator column(s)	Level to be created	Output Sample folder
	Peptide_Id	Intensity B_01	[Intensity A_01 , Intensity A_02 , Intensity A_03 , Intensity A_04] , [Intensity B_01 , Intensity B_02 , Intensity B_03 , Intensity B_04]	u_peptide	b1
	Peptide_Id	Intensity B_02	[Intensity A_01 , Intensity A_02 , Intensity A_03 , Intensity A_04] , [Intensity B_01 , Intensity B_02 , Intensity B_03 , Intensity B_04]	u_peptide	b2
	Peptide_Id	Intensity B_03	[Intensity A_01 , Intensity A_02 , Intensity A_03 , Intensity A_04] , [Intensity B_01 , Intensity B_02 , Intensity B_03 , Intensity B_04]	u_peptide	b3
	Peptide_Id	Intensity B_04	[Intensity A_01 , Intensity A_02 , Intensity A_03 , Intensity A_04] , [Intensity B_01 , Intensity B_02 , Intensity B_03 , Intensity B_04]	u_peptide	b4
	Peptide_Id	Intensity A_01	[Intensity A_01 , Intensity A_02 , Intensity A_03 , Intensity A_04] , [Intensity B_01 , Intensity B_02 , Intensity B_03 , Intensity B_04]	u_peptide	a1
	Peptide_Id	Intensity A_02	[Intensity A_01 , Intensity A_02 , Intensity A_03 , Intensity A_04] , [Intensity B_01 , Intensity B_02 , Intensity B_03 , Intensity B_04]	u_peptide	a2
	Peptide_Id	Intensity A_03	[Intensity A_01 , Intensity A_02 , Intensity A_03 , Intensity A_04] , [Intensity B_01 , Intensity B_02 , Intensity B_03 , Intensity B_04]	u_peptide	a3
	Peptide_Id	Intensity A_04	[Intensity A_01 , Intensity A_02 , Intensity A_03 , Intensity A_04] , [Intensity B_01 , Intensity B_02 , Intensity B_03 , Intensity B_04]	u_peptide	a4

Figure S30. The LEVEL CREATOR task table for workflow 4.

Peptide_Id	Xs_Intensity A_01_vs_Mean_Intensity	Vs_Intensity A_01_vs_Mean_Intensity
6988_EALQSDWLPFELLASGGQK	0.072239228	2378200
6990_EALTYDGALLGDR	0.062201152	144200000
6991_EALVDTLTGILSPVQEVR	0.354203027	43264000
6993_EAMECSDVIWQR	-0.435525774	8966841.667
6998_EAMGIYISTLK	0.106526482	90203000
7000_EAMNDPLLER	0.015057358	43534000

Figure S31. Excerpt from one of the u_peptide files generated by workflow 4 LEVEL CREATOR module showing element identifiers (left column), log2 ratios (center column) and uncalibrated statistical weights (right column).

The *u_peptide* level files are subsequently calibrated using the LEVEL CALIBRATOR module through integration to the protein level (*Figure S32, Top*), resulting in calibrated *peptide* level files. The INTEGRATE module then executes *peptide-to-protein* and *protein-to-proteinall* integrations as specified in the module task table (*Figure S32*, Bottom).

In this example, it is important to note that the advanced option of INTEGRATE was activated to utilize the Tag column, to indicate that only proteins containing the *Homo sapiens* tag are employed in the *protein-to-proteinall* integrations (*Figure S32, Bottom*). The decision to restrict integration to human proteins serves two purposes: a) normalization is performed using the grand mean of human proteins, unaffected by the presence of yeast or E. coli proteins; b) estimation of the variance in the *protein-to-*

proteinall integration relies solely on human proteins, mitigating the impact of yeast and E. coli proteins, which exhibit significant deviations from the mean. Note that this procedure does not remove yeast or E. coli proteins from the normalized files subsequently utilized by the NORCOMBINE module (as explained below).

Sample folder(s)	Lower level for integration	Higher level for integration	Name of calibrated level
*	u_peptide	protein	peptide

Sample folder(s)	Lower level	Higher level	Output Sample folder	Tag	FDR
a1 , a2 , a3 , a4 , b1 , b2 , b3 , b4	peptide	protein			
a1 , a2 , a3 , a4 , b1 , b2 , b3 , b4	protein	proteinall		Homo sapiens	0
B_vs_A	protein	proteinall		Homo sapiens	0

Figure S32. The LEVEL CALIBRATOR (Top) and INTEGRATE (Bottom) task tables for workflow 4.

In this workflow, the protein level comprises numeric values derived from the "Protein group IDs" provided by *ID-q*, generated through the *modificationSpecificPeptides.txt* file from MaxQuant (refer to the *Preparing the ID-q file from MaxQuant output* section). Similar to workflow 3, the *protein-to-proteinall* relation table must include a third column that tags the species from which each protein originates (see *Figure S33*). The species names have been extracted from the *ID-q* file. It is important to note that the tag indicating human proteins aligns with the tag indicated in INTEGRATE (see *Figure S32, Bottom*).

proteinall	protein	Species
1	2469	Homo sapiens
1	2452	Homo sapiens
1	5935	Homo sapiens
1	282	Escherichia coli
1	1279	Saccharomyces cerevisiae
1	5766	Homo sapiens
1	1642	Saccharomyces cerevisiae
1	553	Homo sapiens
1	1007	Homo sapiens
1	1202	Homo sapiens
1	126	Escherichia coli

Figure S33. Excerpt from the protein2proteinall workflow 4 relation table that illustrate the linkage between proteins and a constant value representing the protein grand mean. The protein level is the "Protein group IDs" obtained from ID-q, generated by MaxQuant. It is noteworthy that a third column is employed to tag proteins with their respective species, facilitating later species-specific integration in the protein-to-proteinall context.

The *protein-to-proteinall* relation table is automatically generated by the RELS CREATOR module, which extracts information from the *ID-q* file created by the user. This file includes the relationship

between protein group identifiers (under the *Protein group IDs* column header) and the corresponding species (under the *Species* column header).

Relation Table to be created	Column name of Lower level	Column name of Higher level	Column name of 3rd column	Table from which RT is extracted
u_peptide2protein	Peptide_Id	Protein group IDs		
peptide2protein	Peptide_Id	Protein group IDs		
protein2proteinall	Protein group IDs	[1]	Species	

Figure S34. The RELS CREATOR task table for workflow 4.

Next, the normalized data at the *protein* level from the four replicates from each sample are combined into samples *A* and *B*, respectively, using the NORCOMBINE basic module (*Figure S35*, Top). To compare these two samples, new log2 ratios and statistical weights are calculated using the RATIOS basic module (*Figure S35*, Bottom). Finally, a *protein-to-proteinall* integration is carried out for the newly generated *B_vs_A* sample by the module INTEGRATE (*Figure S32*, Bottom), using again the *Homo sapiens* tag.

Sample folders	Level	Norm	lowerNorm	Output Sample folder
b1 , b2 , b3 , b4	protein	proteinall	lowerNormV	B
a1 , a2 , a3 , a4	protein	proteinall	lowerNormV	A

Numerator Sample folder	Denominator Sample folder(s)	Level	V Method	Output Sample folder
B	A	protein	avg	B_vs_A

Figure S35. The NORCOMBINE (Top) and RATIOS (Bottom) task tables for workflow 4.

The REPORT module, employed in this workflow, compiles data at the protein level alongside the number of peptides per protein, like previous workflows (*Figure S36*). In this specific instance, the table encompasses Z and FDR, as well as the log2ratios of proteins from all samples (*Xinf* from the *protein-to-proteinall* integration, also denoted as *Xq*), the grand mean (*Xsup* from the *protein-to-proteinall* integration, or *Xa*), the statistical weights (*Vinf* or *Vq*), and the respective species (*tags*) for each protein. The grand mean, utilized for log2-ratio normalization, along with the statistical weights, can be employed in constructing plots, as illustrated in *Figure S37* (see below).

Sample folder(s)	Lower level	Higher level	Reported vars	Output report	Column headers to eliminate	Merge with report
*	peptide	protein	n	Npep2prot		
*	protein	proteinall	Xinf, Xsup, Vinf , Z , FDR , tags	Npep2prot_Quanprot	peptide	Npep2prot

Figure S36. The REPORT module task table for workflow 4.

It is noteworthy that variance modelling, normalization, standardization and statistical weighting, according to the GIA algorithm, are performed automatically, without data filtering, pre-processing or

missing value imputation [6], even in a situation where numerous proteins have highly imbalanced data. Moreover, the sigmoid plots automatically generated in each one of the integrations performed by INTEGRATE clearly demonstrate that the GIA algorithm accurately predicts the distribution of peptide quantifications around their proteins (*Figure S37A*) and of protein quantifications around the grand mean (*Figure S37B*). These results demonstrate that this statistical model is very suitable for the analysis of label-free data.

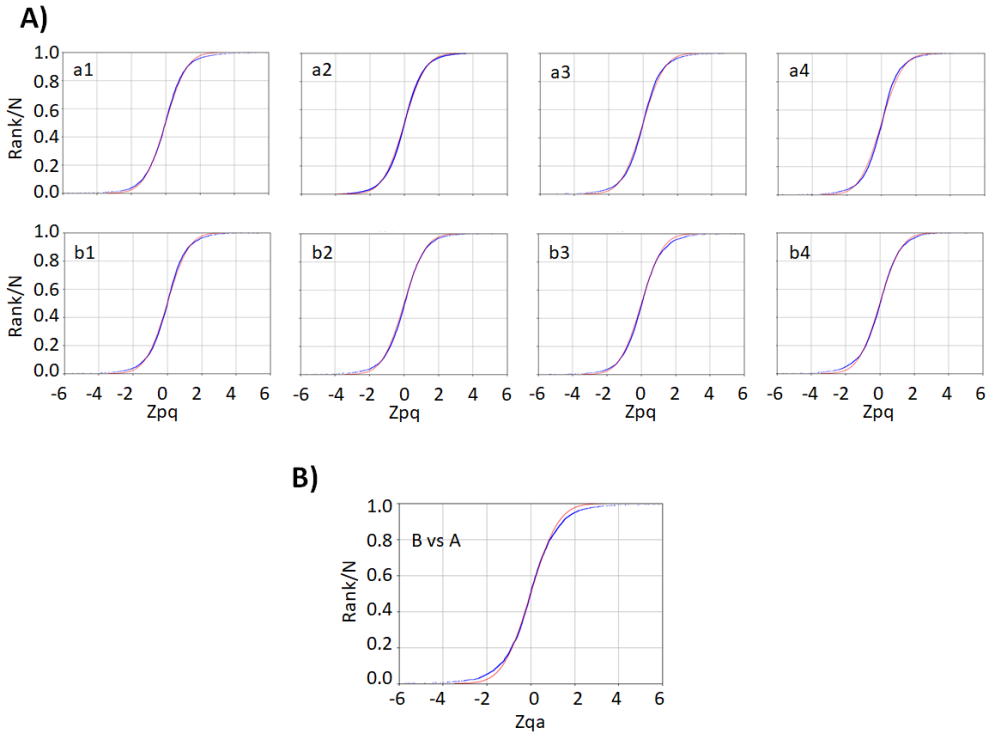


Figure S37. Distribution of the standardized variable at the peptide (Zpq) and protein (Zqa) levels for label-free data analyzed with iSanXoT. A) Zpq distribution for the eight individual A-type and B-type samples. B) Zqa distribution for the B-type vs A-type comparison. Red: null hypothesis (standard distribution); blue: experimental data.

The combined statistics *B_vs_A* also shows how human, yeast and bacterial proteins distribute approximately around the expected 0, 1, and -2 log2-values (corresponding to 1-, 2- and 0.25-fold changes) (*Figure S38*), and how protein quantifications with higher statistical weights are more accurate. This plot also confirms how iSanXoT provides highly accurate quantitative results in a fully automated fashion.

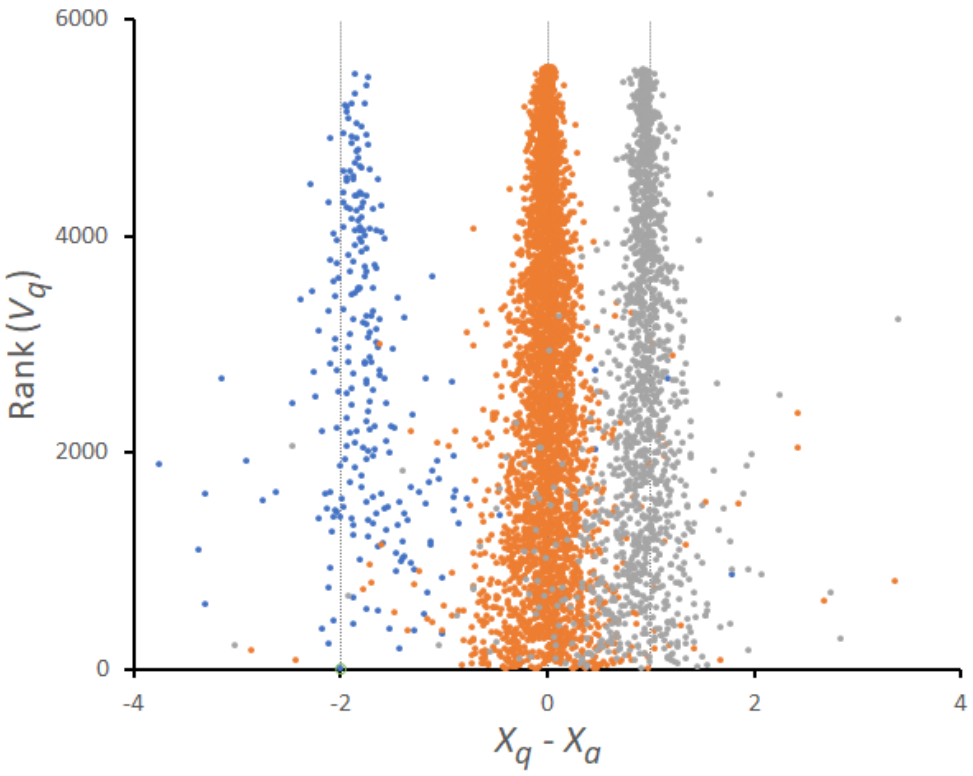


Figure S38. Quantification of human (orange), yeast (grey) and bacterial (blue) proteins according to the combined statistics B_vs_A. Shown are log2-ratios normalized by the grand mean (Xq – Xa or Xinf - Xsup). This plot was generated from the table "Npep2prot_Quanprot".

As a means to benchmark the performance of iSanXoT for label-free data, we counted up how many proteins exhibited statistically significant changes between the two preparations. Proteins originating from E. coli and yeast were classified as true positives, while those from *Homo sapiens* were labeled as false positives. Subsequently, we computed key metrics, including the False Positive Rate (FPR), True Positive Rate (TPR), and False Discovery Rate (FDR), based on these classifications.

The metrics were applied following two distinct approaches. In the first approach, the integrated results of each sample for each condition (a1, a2, a3, a4 and b1, b2, b3, b4) were considered separately to obtain a standardized value for each protein (Zqa). Subsequently, a conventional t-test was conducted to compare Zqa values between the two samples and the *p-values* were adjusted for multiple hypothesis testing (FDR <= 0.05). The results are presented in Table S2.

Table S2. Computation of False Positive Rate (FPR), True Positive Rate (TPR), and False Discovery Rate (FDR) based on statistically significant proteins from the integrated results of samples for each

condition provided by iSanXoT and the conventional t-test. The FPR, TPR, and FDR percentages were calculated from the quantified proteins of iSanXoT and the identified proteins of the search engine.

					from quantified proteins			from identified proteins		
iSanXoT (t-test)		**identified proteins/protein group**	**quantified proteins/protein groups**	**FDR <= 0.05**	**FPR**	**TPR**	**FDR**	**FPR**	**TPR**	**FDR**
	Escherichia coli	433	196	185		94.4%			42.7%	
	Homo sapiens	3689	3286	141	4.3%		10.6%	3.8%		2.5%
	Saccharomyces cerevisiae	1519	1183	1005		85.0%			66.2%	
	Total	5641	4665	1331						

Note that since iSanXoT does not input missing values and the t-test cannot be calculated in cases when there are fewer than two replicates, not all identified proteins could be quantified (*Table S2*). In the subset of quantified proteins, iSanXoT got TPRs near 90% while maintaining the FPR below 5%.

In the second approach, we directly calculated the statistically significant protein changes from the combined statistics *B_vs_A*. This was done counting up how many Zqa values significantly deviated from the expected N(0.1) distribution at FDR <= 0.05.

Table S3. Computation of False Positive Rate (FPR), True Positive Rate (TPR), and False Discovery Rate (FDR) based on significant changes determined by the iSanXoT modules (NORCOMBINE, RATIO, INTEGRATE). The FPR, TPR, and FDR percentages were calculated from the quantified proteins of iSanXoT and the identified proteins of the search engine.

					from quantified proteins			from identified proteins		
iSanXoT (NORCOMBINE)		**identified proteins/protein group**	**quantified proteins/protein groups**	**FDR <= 0.05**	**FPR**	**TPR**	**FDR**	**FPR**	**TPR**	**FDR**
	Escherichia coli	433	253	248		98.0%			57.3%	
	Homo sapiens	3689	3643	218	6.0%		12.7%	5.9%		3.9%
	Saccharomyces cerevisiae	1519	1380	1249		90.5%			82.2%	
	Total	5641	5276	1715						

As shown in *Table S3*, iSanXoT was able to quantify more proteins than in the previous case. This was possible without performing missing value imputation because the statistical model assigned a specific variance to all the proteins integrated by NORCOMBINE, independently of the number of replicates where the protein was quantified in each of the samples. Note that all the protein variances were corrected by the variance detected by the model at the time of averaging replicates.

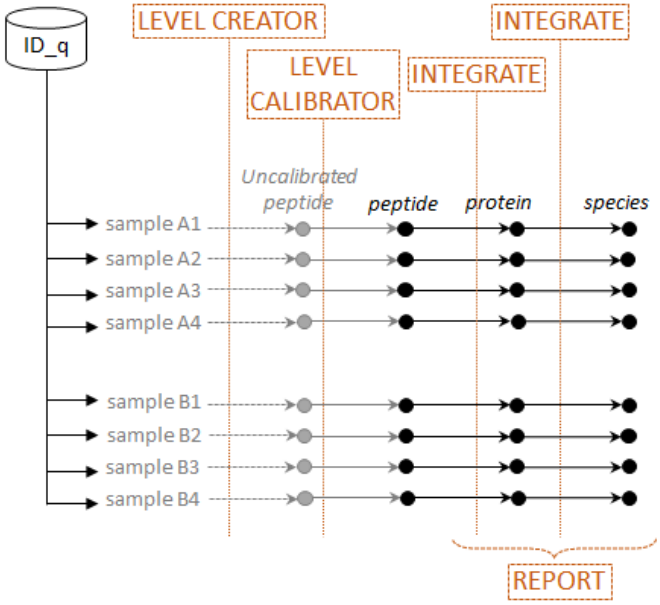
The versatility of iSanXoT allows for the creation of unlimited integrations depending on the desired levels. In this specific experiment, where the proteins come from several species, we can integrate the protein values to the species they belong to. This integration normalizes the protein data within each species, producing standardized protein values that describe the deviation from each species' average. To perform this integration, we have to define a protein2species relation table using the RELS_CREATOR module (*Figure S39*).

Relation Table to be created	Column name of Lower level	Column name of Higher level	Column name of 3rd column
u_peptide2protein	Peptide_Id	Protein group IDs	
peptide2protein	Peptide_Id	Protein group IDs	
protein2proteinall	Protein group IDs	[1]	Species
protein2species	Protein group IDs	Species	

Figure S39. RELS_CREATOR module for the protein-to-species workflow.

Again, we can follow two approaches. In the first one each sample is separately integrated to species to obtain a standardized Zqs value for each protein (note that in this case the *s* subscript refers to the species level). A conventional t-test can then be applied to the set of Zqs values (*Figure S40A*), to detect whether there are significant differences between the two samples. Here a statistically significant change would indicate that the protein deviates from the rest of proteins of the same species. To construct these integrations, we only need to indicate protein and species as the lower and higher levels, respectively in the INTEGRATE module (*Figure S40B*).

A)



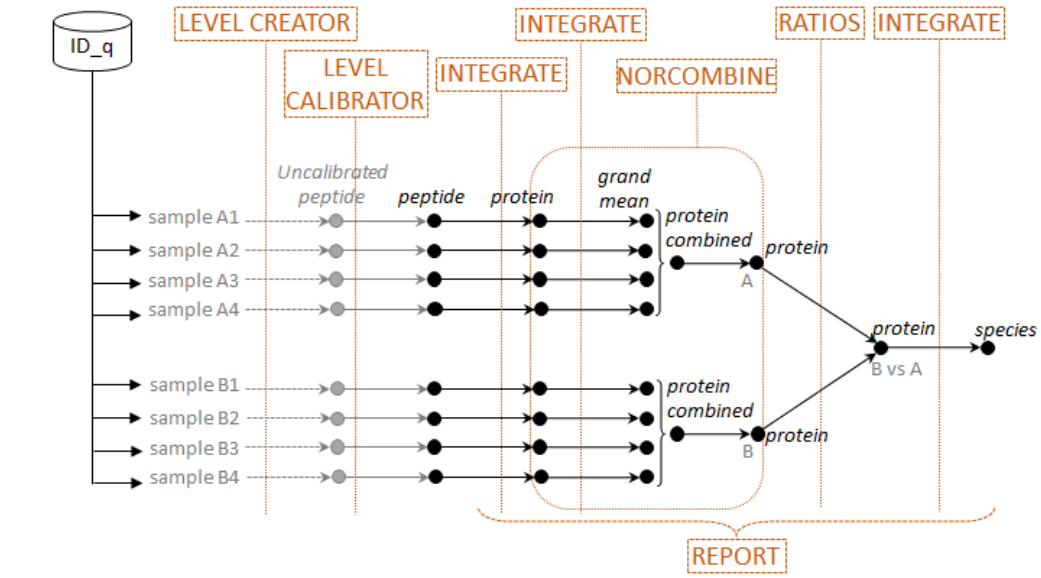
B)

Sample folder(s)	Lower level	Higher level	Output	Tag	FDR
a1 , a2 , a3 , a4 , b1 , b2 , b3 , b4	peptide	protein			
a1 , a2 , a3 , a4 , b1 , b2 , b3 , b4	protein	species			0

Figure S40. (A) Scheme of the protein-to-species workflow (label-free quantification), integrating each sample separately. (B) The INTEGRATE module that allows integration among the levels described in (A). The outliers in the protein-to-species analysis are not removed (set to 0 for FDR).

In the second approach, the protein values in each sample are integrated using NORCOMBINE and the ratio between the samples calculated using RATIO, as in *Figure S35*. But now the resulting level is integrated to species (*Figure S41A*). The statistically significant protein changes are then detected from the Zqa values in the *B_vs_A* statistics, as before. To perform this, the INTEGRATE task table requires a new *protein-to-species* integration specifically for the *B_vs_A* combined statistic sample (*Figure S41B*).

A)



B)

Sample folder(s)	Lower level	Higher level	Output Sample folder	Tag	FDR
a1 , a2 , a3 , a4 , b1 , b2 , b3 , b4	peptide	protein			
a1 , a2 , a3 , a4 , b1 , b2 , b3 , b4	protein	proteinall		Homo sapiens	0
B_vs_A	protein	species			0

Figure S41. Schematic representation of the protein-to-species workflow, involving NORCOMBINE and RATIO modules for processing protein values. (B) The INTEGRATE module facilitating integration across the levels described in (A).

The REPORT module utilized in the *protein-to-species* workflows (*Figure S42*), resembles workflow 4 (*Figure S36*), with the distinction of exclusively incorporating the *protein-to-species* integration to derive the necessary values for the analysis.

Sample folder(s)	Lower level	Higher level	Reported vars	Output report	Column headers to eliminate	Merge with report
*	peptide	protein	n	Npep2prot		
*	protein	proteinall	Xinf, Xsup, Vinf , Z , FDR , tags	Npep2prot_Quanprot	peptide	Npep2prot
*	protein	species	Xinf, Xsup, Vinf , Z , FDR , tags	Npep2prot_Quanprot_Species	peptide	

Figure S42. The REPORT module task table for protein-to-species workflows.

As mentioned earlier, the coordinated behavior of proteins within each species can be analyzed by inspecting the automatically-generated sigmoid plots (*Figure S43A*), clearly illustrating that the GIA algorithm accurately predicts the distribution of protein quantifications. Similarly, the combined *B_vs_A* statistics, when integrated to species, demonstrate how human, yeast, and E. coli proteins distribute approximately around the expected log2-values of 0 (*Figure S43B*).

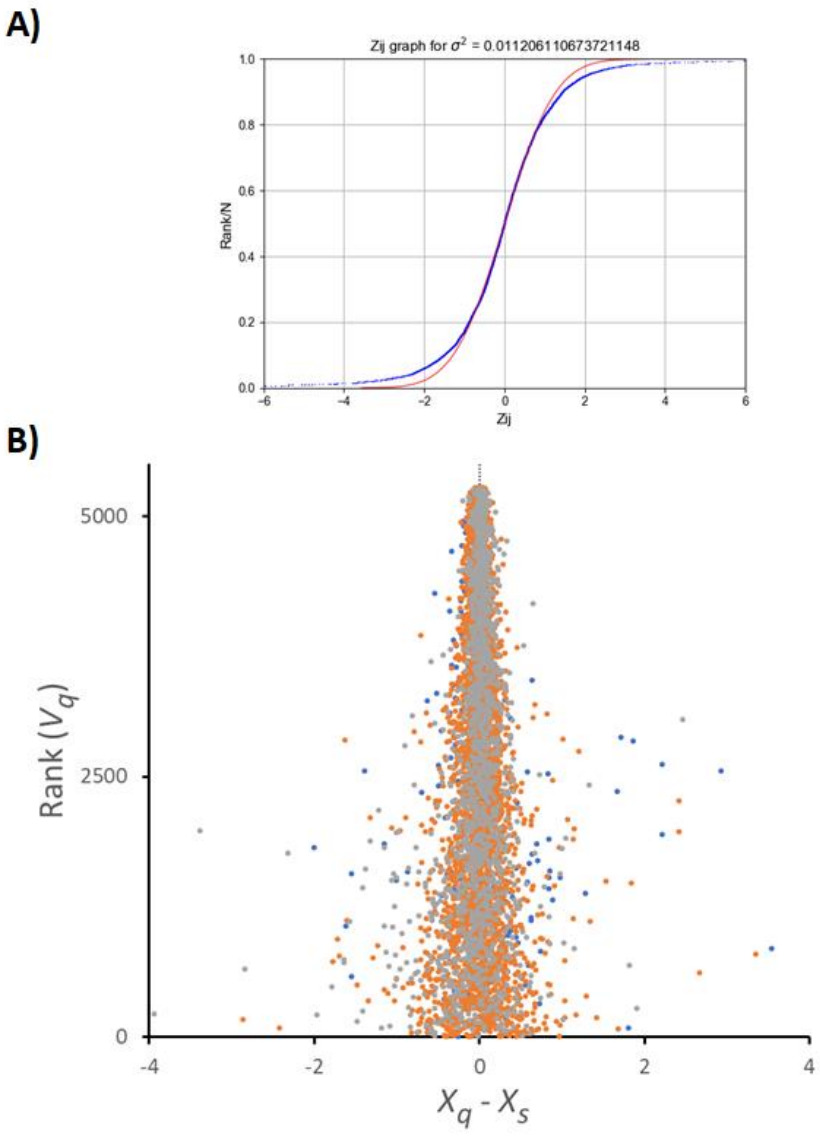


Figure S43. (A) Distribution of the standardized log2 protein ratios (Zqs) resulting from the protein-to-species integration of B_vs_A. Red represents the null hypothesis (standard distribution), and blue represents the experimental data. (B) Quantification of human (orange), yeast (grey), and E. coli (blue) proteins using the protein-to-species workflow with NORCOMBINE. The log2-ratios are normalized by the respective species (Xq – Xs). The letters 'q' and 's' denote the protein and species, respectively.

The performance of iSanXoT in the two approaches are summarized in Table S4. Here, the null hypothesis is that each protein follows the same quantitative behaviour as the rest of proteins of the same species. Hence, we can calculate a FPR per species. As explained above, iSanXoT does not handle missing values as input, and hence the first approach (*Table S4A*) quantifies less proteins than the second (*Table S4B*). Similarly, the second test is more sensitive to detect outliers. In both cases, human and yeast proteins are quantified with low FPRs. In contrast, this analysis highlights how E.coli proteins are more difficult to quantify due to the large difference in abundance in the original samples.

Table S4. Computation of False Positive Rate (FPR), based on the significant proteins obtained from the protein-to-species workflow and the conventional t-test (A) and from the iSanXoT modules (NORCOMBINE, RATIO, INTEGRATE) (B). The FPR values were calculated from the quantified proteins of iSanXoT.

A)

		Analysis at the species level (significant changes)			from quantified proteins
iSanXoT (t-test)		**identified proteins/protein group**	**quantified proteins/protein groups**	**FDR <= 0.05**	**FPR**
	Escherichia coli	433	196	6	3.1%
	Homo sapiens	3689	3286	12	0.4%
	Saccharomyces cerevisiae	1519	1184	5	0.4%
	Total	5641	4666	23	

B)

		Analysis at the species level (significant changes)			from quantified proteins
iSanXoT (NORCOMBINE)		**identified proteins/protein group**	**quantified proteins/protein groups**	**FDR <= 0.05**	**FPR**
	Escherichia coli	433	253	43	17.0%
	Homo sapiens	3689	3643	105	2.9%
	Saccharomyces cerevisiae	1519	1380	77	5.6%
	Total	5641	5276	225	

Importing a workflow template

In this section, we will provide instructions to execute the workflow examples and to import workflows that were previously created with iSanXoT to be reused in other projects. We will use the first workflow described in the previous section as an example.

Start by downloading the template for Workflow 1 and the input files from the iSanXoT documentation (https://raw.githubusercontent.com/CNIC-Proteomics/iSanXoT/master/docs/templates/WSPP-SBT.zip).

Then, extract the files included in the compressed archive to create a folder named WSPP-SBT. Check that the WSPP-SBT folder has been created in your file system. Proceed as follows:

- Open the iSanXoT application by double-clicking the application icon (*Figure S44*).

Figure S44. The iSanXoT startup message.

- Choose *New Project* from the *Project* menu (*Figure S45*).

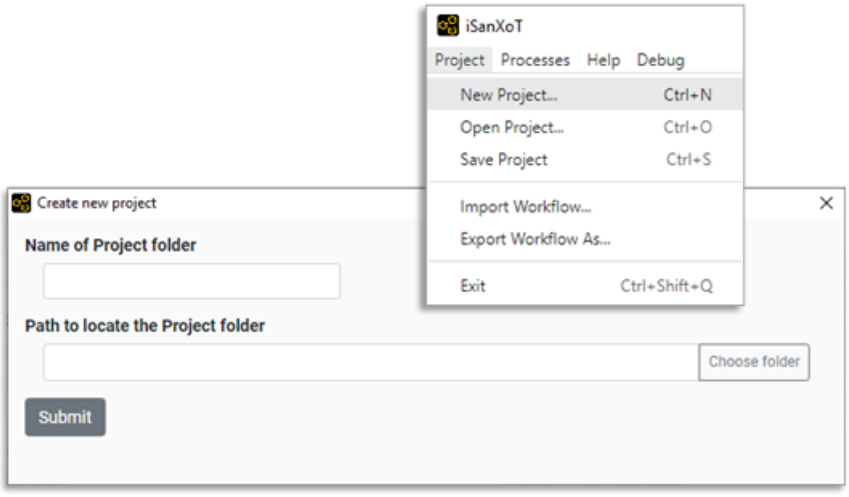


Figure S45. Create New Project.

- Provide a name of your choice for the project folder and indicate a path to locate this folder, then click the *Submit* button (*Figure S45*).

- Choose *Import Workflow* from the *Project* menu (*Figure S46)* and select the folder WSPP-SBT created before (or any other iSanXoT project folder from which you want to import the workflow).

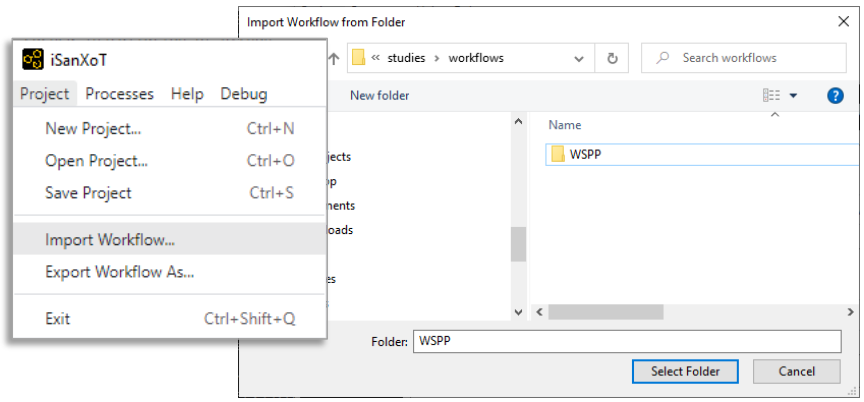


Figure S46. Importing a preexisting iSanXoT workflow to the newly-created project.

- Inspect the WSPP-SBT task table (in the Compound modules tab), the RELS CREATOR task table (in the Relation tables tab) and the REPORT task table (in the Reports tab) to check that the tables indicated in Fig. S2, S3 and S5 have been correctly loaded. Note that if a different template is imported, only the corresponding task tables will be loaded.

- Now, click on *Choose identification file* and select "ID-q.tsv" in the WSPP-SBT folder (*Figure S47)*. Alternatively, select the desired identification/quantification table with which this workflow is to be executed. Section 3 below shows how to prepare the "ID-q" file based on the output from a variety of proteomics pipelines. Bear in mind that the tasks defined in the LEVEL CREATOR and RELS CREATOR modules have to match the samples and column names from the specific "ID-q" file used.

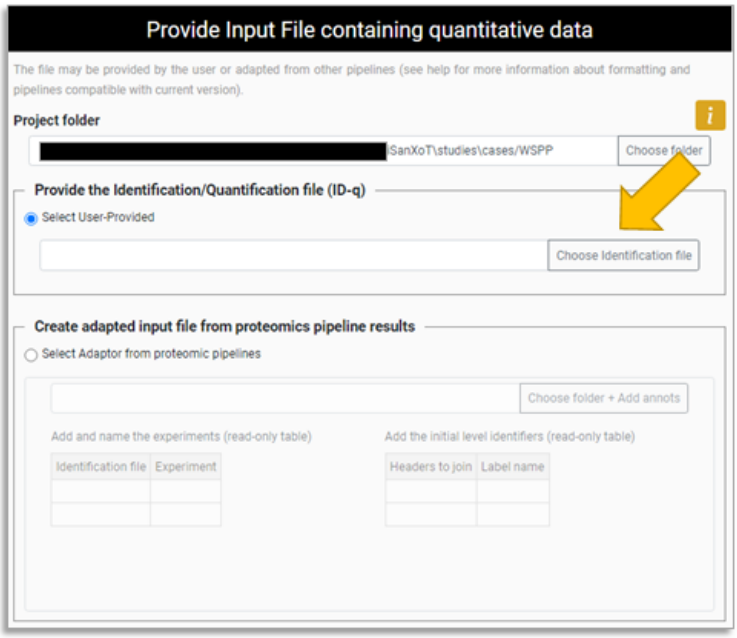


Figure S47. Choosing the identification/quantification (ID-q) file for the newly-created project.

- Select *Save Project* from the *Project* menu to save the changes, or directly press the *Save and Run* button to both save and execute the current workflow.

Creating the identification/quantification file from proteomics pipelines

iSanXoT necessitates an identification/quantification file in TSV format (*ID-q.tsv*), which should include, at a minimum, the quantified features alongside their corresponding quantitative values. Any TSV table can serve as an ID-q file, given that quantitative values are organized with column headers. In this arrangement, features (such as PSMs or peptides) are listed in rows, and their quantitative values are presented in columns, with each column representing a distinct sample. iSanXoT utilizes the column headers of the *ID-q* file to extract essential information.

Moreover, when the *ID-q* file encompasses features quantified in more than one experiment (e.g., different samples labeled with the same TMT-18plex tags), it must include an additional column with the header "Experiment" to specify the experiment assignment of the features.

Finally, iSanXoT requires additional information to generate the relation files necessary for integrating quantified features into higher levels. This information is typically found in the *ID-q* file. For example, iSanXoT can utilize columns containing scan and peptide identifiers to construct the *scan2peptide* relation table.

The majority of proteomics software tools generate tables that can be easily used for this purpose. In this section, we will describe how to prepare the *ID-q* file based on the output from the three most popular proteomics pipelines (*Table S5*).

Table S5. Output data from proteomics pipelines to be included in the ID-q.tsv *file.*

Proteomics pipeline	Experiment type	Output file name / suffix	Level name in the output file			Quantitative data
			Scan	Peptide	Protein	
Proteome Discoverer (version 2.5)	Label-free	*_PeptideGroups.txt*		*Sequence + Modifications*	*Master Protein Accessions*	*Abundance: FX: Sample Type*
	Isotopically labelled	*_PSMs.txt*	*Spectrum File + First Scan*[1]	*Sequence + Modifications*	*Master Protein Accessions*	*Abundance: Quan Channel*
MaxQuant (version 1.6.5.0)	Label-free	*modificationSpecificPeptides.txt*		*Sequence + Modifications*	*Proteins*	*Intensity Experiment*
	Isotopically labelled	*msmsScans.txt*	*Raw file + Scan number*	*Modified Sequence*	*Proteins*	*Reporter intensity n*
Fragpipe (version 1.8.1)	Label-free	*combined_modified_peptide.tsv*		*Modified Sequence*	*Protein ID*	*Experiment Intensity*
	Isotopically labelled	*psm.tsv*	*Spectrum + Spectrum File*[2]	*Modified Peptide*	*Protein ID*	*Channel*

[1]Make sure *Max. Number of Peptides Reported* = 1 was selected in the *Input Data* section of the Proteome Discoverer *Processing* node used.

[2]Make sure *Report top N* = 1 was selected in the *Advanced Output Options* of the FragPipe *MSFragger* module.

Preparing the *ID-q* file from Proteome Discoverer output

In the case of Proteome Discoverer version 2.5 [8], the way that quantitative data are adapted for use with iSanXoT depends on whether they originate from label-free or labelled experiments:

Label-free experiments

In this scenario, quantitative data at the peptide level can be adapted for use with iSanXoT from the *_PeptideGroups.txt* files obtained when the *Processing* workflow node *Minora Feature Detector* of Proteome Discoverer is employed. The following column headers of the *_PeptideGroups.txt* files must be considered when preparing the *ID-q* file:

- *Sequence*: Amino acid sequence of the identified peptide;
- *Modifications*: Chemical or posttranslational modifications to the *Sequence* above;
- *Master Protein Accessions*: Accession code(s) for the protein(s) to which the peptide *Sequence* is ascribed;
- *Abundance: FX: Sample Type:* Peptide intensity in the RAW file identified with *FX* and tagged as *Sample Type* in the Proteome Discoverer *Input Files* tab.

The *peptide* level required for the *peptide-to-protein* integration with iSanXoT can be obtained by merging the *Sequence* and *Modifications fields* (see Section *Adapting the results from proteomics pipelines for iSanXoT* below).

Labelled experiments

For labelled experiments (e.g., TMT- or iTRAQ-based), quantitative data at the scan level can be adapted for use with iSanXot from the *_PSMs.txt* files generated when the *Processing* workflow node *Reporter Ions Quantifier* of Proteome Discoverer is used. The following column headers of the *_PSMs.txt* files must be considered for preparing the *ID-q* file:

- *Spectrum File*: Name of the RAW file where the PSM was identified;
- *First Scan*: Spectrum (scan) number of the PSM in the RAW file;
- *Sequence*: Amino acid sequence of the identified peptide;
- *Modifications*: Chemical or posttranslational modifications to the *Sequence* above;
- *Master Protein Accessions*: Accession code(s) for the protein(s) to which the peptide *Sequence* is ascribed;
- *Abundance: Quan Channel*: Intensity of the reporter ion tagged as *Quan Channel* in the Proteome Discoverer *Samples* tab.

For the *scan to peptide* integration with iSanXoT, the *scan* level can be obtained by merging the *Spectrum File* and *First Scan* fields, and the *peptide* level by merging the *Sequence* and *Modifications* fields (see Section *Adapting the results from proteomics pipelines for iSanXoT* below; make sure *Max. Number of Peptides Reported* = 1 was selected in the *Input Data* section of the Proteome Discoverer *Processing* node used).

Preparing the *ID-q* file from MaxQuant output

The way MaxQuant version 1.6.5.0 [7] data are adapted for use with iSanXoT depends on whether they originate from label-free or labelled proteomics experiments:

Label-free experiments

In this case, the quantifications at the peptide level required to prepare the *ID-q* file can be found in the *modificationSpecificPeptides.txt* file, which is stored in the *"…combined/txt*" folder. The following column headers of the *modificationSpecificPeptides.txt* file must be considered for preparing the *ID-q* file:

- *Sequence*: Amino acid sequence of the identified peptide;
- *Modifications*: Chemical or posttranslational modifications to the *Sequence* above;
- *Proteins*: Identifier(s) of the protein(s) to which the peptide *Sequence* is ascribed;
- *Intensity Experiment*: Summed up extracted ion current of all isotopic clusters associated with the peptide Sequence identified across the raw files included in the *Experiment* as specified by the user in the MaxQuant *Raw data* tab.

The *peptide* level required for the *peptide* to protein integration with iSanXoT can be obtained by merging *Sequence* and *Modifications* fields (see Section *Adapting the results from proteomics pipelines for iSanXoT* below).

Labelled experiments

When dealing with labelled experiments (e.g., iTRAQ- or TMT-based), the necessary quantitative data at the scan level can be found in the *modificationSpecificPeptides.txt* file, which is stored in the *"…combined/txt*" folder. The following column headers of the *modificationSpecificPeptides.txt* file must be considered for preparing the *ID-q* file:

- *Raw file*: Name of the RAW file where the PSM was identified;
- *Scan number*: Spectrum (scan) number of the PSM in the RAW file;
- *Modified Sequence*: Amino acid sequence of the identified peptide including chemical or posttranslational modifications. This parameter is nonblank only when identification was successful.
- *Proteins*: Identifier(s) of the protein(s) to which the peptide *Sequence* is ascribed;
- *Reporter intensity n*: Intensity of the reporter ion *n* as specified by the user in the MaxQuant *Group-specific parameters* tab.

For the *scan to peptide* integration with iSanXoT, the *scan* level can be obtained by merging the *Raw File* and *Scan number* fields (see Section *Adapting the results from proteomics pipelines for iSanXoT* below).

Preparing the *ID-q* file from FragPipe output

The way that quantitative data from Fragpipe version 1.8.1 [9] are adapted for use with iSanXoT depends on whether they originate from label-free or labelled experiments:

Label-free experiments

FragPipe *Quant (MS1)* module stores the quantifications at the peptide level necessary to prepare the *ID-q* file in a *combined_modified_peptide.tsv* file. The following column headers of the *modificationSpecificPeptides.txt* file must be considered for preparing the *ID-q* file:

- *Modified Sequence*: Amino acid sequence of the identified peptide;
- *Protein ID*: Identifier of the protein to which the *Modified Sequence* peptide is ascribed;
- *Experiment Intensity*: Summed up intensity of the *Modified Sequence* peptide in the RAW files included in the *Experiment* as specified by the user in the FragPipe *Workflow* tab.

Labelled experiments

Fragpipe *Quant (Isobaric)* module generates a *psm.tsv* output file that contains the quantitative data at the scan level obtained from labelled experiments. The following column headers of the *psm.txt* file must be considered for preparing the *ID-q* file:

- *Spectrum*: Spectrum (scan) identifier of the PSM in the XML file;
- *Spectrum File*: Name of the XML file where the PSM was identified;
- *Modified Peptide*: Amino acid sequence of the identified peptide including chemical or posttranslational modifications;
- *Protein ID:* Identifier of the protein to which the *Modified Peptide* is ascribed;
- *Channel*: Intensity of the reporter ion *Channel* as specified by the user in the FragPipe TMT-Integrator table of the *Quant (Isobaric)* module.

The *scan* level required for the later *scan to peptide* integration with iSanXoT can be obtained by merging the *Spectrum* and *Spectrum File* fields (see Section *Adapting the results from proteomics pipelines for iSanXoT* below; make sure *Report top N = 1* was selected in the *Advanced Output Options* of the FragPipe *MSFragger* module).

Adapting the results from proteomics pipelines for iSanXoT

iSanXoT requires an identification/quantification tab-separated values file (*ID-q.tsv*) containing at least the identified features along with their quantitative values (an experiment identifier is required if two or more experiments are included). Users can either manually compose this *ID-q* file (refer to the previous section for guidance on how to do this using data from the four most popular proteomics pipelines) or have it prepared by the iSanXoT Input Adaptor. The latter option is described in this section.

- Run the iSanXoT application and create a new project or open an existing project. A new window will appear asking for the ID-q file (*Figure S48*).

- If you already have a suitable *ID-q* file, click the *Select User-Provided* option, and then *Choose identification file* to select the file (*Figure S48*).

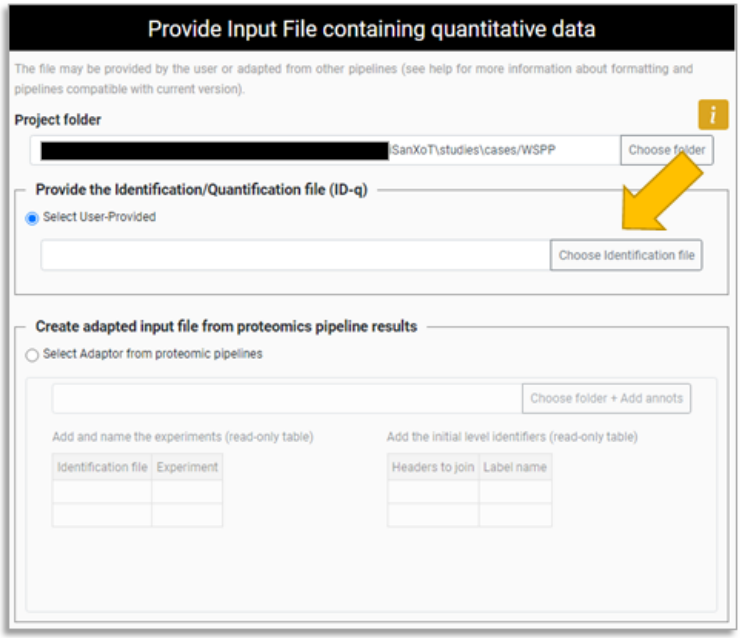


Figure S48. Selecting an ID-q file in the Input Adaptor main window.

- If you do not have an *ID-q* file, click *Select Adaptor from proteomics pipelines*. This option will launch the iSanXoT adapter to import your quantitative data. The adapter has been tested with recent versions of MaxQuant, Trans-Proteomic Pipeline, FragPipe, and Proteome Discoverer. Click on *Choose folder + Add anots* to select the folder containing your quantitative data (*Figure S49*). A three-panel window will pop-up (*Figure S50*).

Figure S49. Having the iSanXoT Input Adaptor prepare the ID-q file.

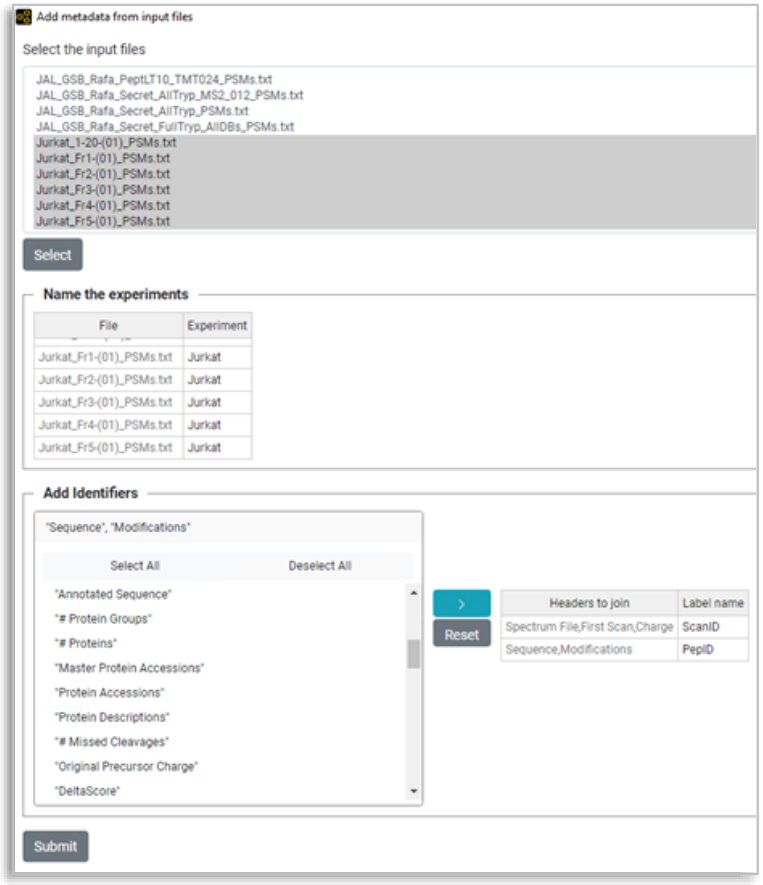


Figure S50. Adapting results from a proteomics pipeline. In the top panel, several output files from Proteome Discoverer have been selected. These PSMs.txt files, which contain identification/quantification data, have been assigned an experiment name (Jurkat) in the middle panel. The bottom panel has been used to create identifiers by concatenating result file headers: ScanID (by concatenating Spectrum File, First Scan and Charge) and pepID (by concatenating Sequence and Modifications).

- The top panel displays the files included in the folder, allowing you to select one or more result files for consideration by the adapter. It's important to note that if several result files are chosen, they must have the same column headers.

- The middle panel is used to set the distribution of data items across experiments according to result filenames.

- The bottom panel allows to create identifiers by concatenating result file headers. It is composed of two interfaces:
    - The left side lists the headers found in the result files, with selected header names being added to the right-side interface.
    - The right-side interface displays the selected header names used to generate the identifier, along with the user-provided identifier name.

- Please note that the alphanumeric text that unambiguously identifies the items to be integrated is the only identifier that must be necessarily included in the *ID-q* file.

- Click the *Submit* button and the Input Adaptor will start generating the *ID-q.tsv* file.

References

[1] Garcia-Marques F, Trevisan-Herraz M, Martinez-Martinez S, Camafeita E, Jorge I, Lopez JA, et al. A Novel Systems-Biology Algorithm for the Analysis of Coordinated Protein Responses Using Quantitative Proteomics. Mol Cell Proteomics. 2016;15(5):1740-60.

[2] Navarro P, Trevisan-Herraz M, Bonzon-Kulichenko E, Nunez E, Martinez-Acedo P, Perez-Hernandez D, et al. General statistical framework for quantitative proteomics by stable isotope labeling. J Proteome Res. 2014;13(3):1234-47.

[3] Gonzalez-Amor M, Garcia-Redondo AB, Jorge I, Zalba G, Becares M, Ruiz-Rodriguez MJ, et al. Interferon stimulated gene 15 pathway is a novel mediator of endothelial dysfunction and aneurysms development in angiotensin II infused mice through increased oxidative stress. Cardiovasc Res. 2021.

[4] Bonzon-Kulichenko E, Camafeita E, Lopez JA, Gomez-Serrano M, Jorge I, Calvo E, et al. Improved integrative analysis of the thiol redox proteome using filter-aided sample preparation. J Proteomics. 2020;214:103624.

[5] Bagwan N, Bonzon-Kulichenko E, Calvo E, Lechuga-Vieco AV, Michalakopoulos S, Trevisan-Herraz M, et al. Comprehensive Quantification of the Modified Proteome Reveals Oxidative Heart Damage in Mitochondrial Heteroplasmy. Cell Rep. 2018;23(12):3685-97 e4.

[6] Navarro P, Kuharev J, Gillet LC, Bernhardt OM, MacLean B, Rost HL, et al. A multicenter study benchmarks software tools for label-free proteome quantification. Nat Biotechnol. 2016;34(11):1130-6.

[7] Tyanova S, Temu T, Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat Protoc. 2016;11(12):2301-19.

[8] Orsburn BC. Proteome Discoverer-A Community Enhanced Data Processing Suite for Protein Informatics. Proteomes. 2021;9(1).

[9] Kong AT, Leprevost FV, Avtonomov DM, Mellacheruvu D, Nesvizhskii AI. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. Nat Methods. 2017;14(5):513-20.