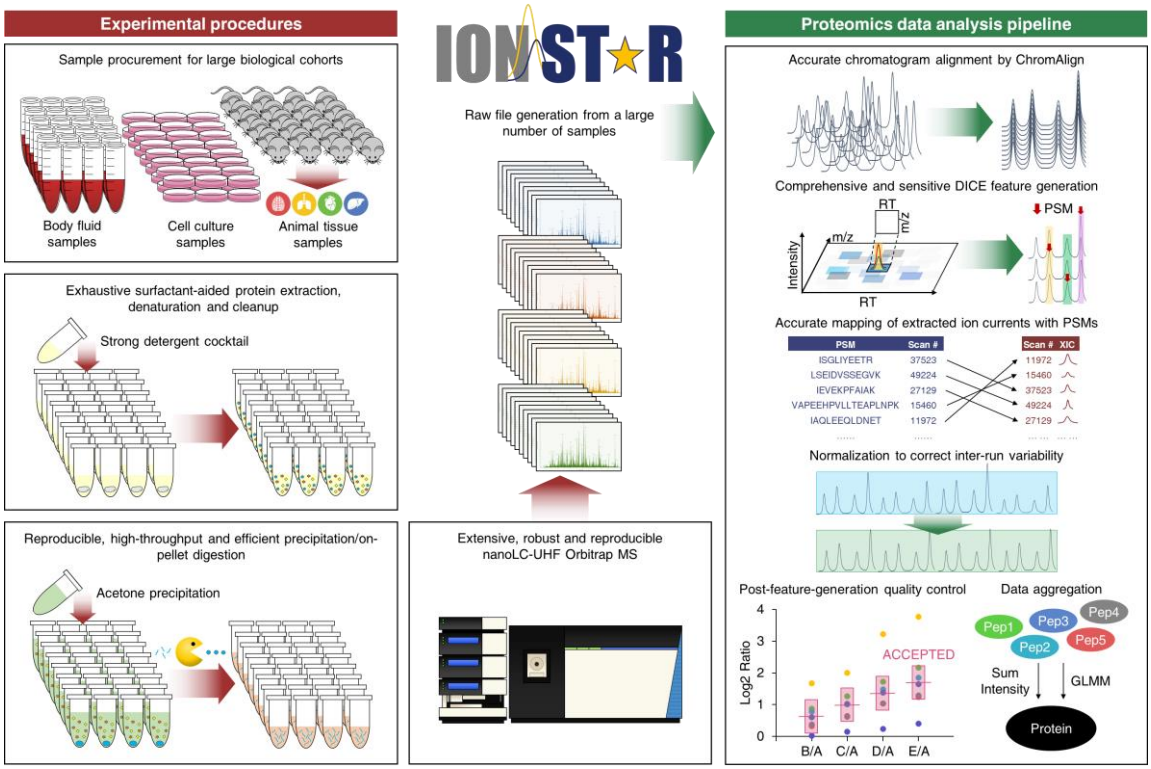
USER MANUAL for Build v0.1.4

TABLE OF CONTENTS

INTRODUCTION

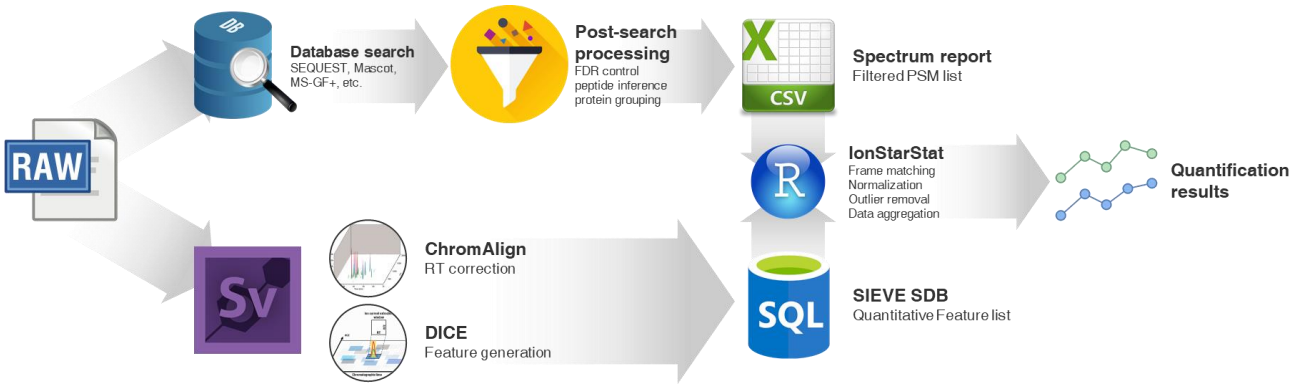


IonStar is an MS1-based quantitative method for label-free proteomics experiments, devised to address issues related with quantitative precision, missing data, and false-positive discovery of protein changes in large-cohort analysis. IonStar comprises of two parts: experimental procedures (left panel) and a proteomics data analysis pipeline (right panel). Details of the experimental procedures can be found in the following literature:

- Shen X, Shen S, Li J, Hu Q, Nie L, Tu C, Wang X, Orsburn B, Wang J, Qu J*. An IonStar experimental strategy for MS1 ion current-based quantification using ultra-high-field Orbitrap: reproducible, in-depth and accurate protein measurement in larger cohorts. *J Proteome Res.* 16(7):2445-2456. (2017)
- An B, Zhang M, Johnson RW, Qu J*. Surfactant-Aided Precipitation/On-Pellet-Digestion (SOD) Procedure Provides Robust and Rapid Sample Preparation for Reproducible, Accurate and Sensitive LC-MS Quantification of Therapeutic Protein in Plasma and Tissues. *Anal Chem.* 87(7):4023-9 (2015)

This manual will focus on the data analysis pipeline part of IonStar, including the functional modules, overall workflow, and the detailed usage of each module. An empirical example will also be provided to help IonStar users to run the pipeline in their own environment.

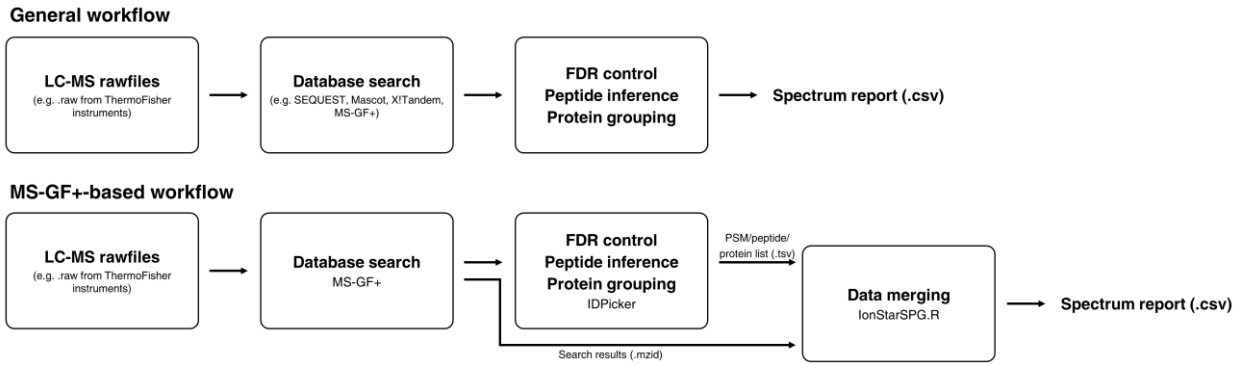
OVERALL WORKFLOW



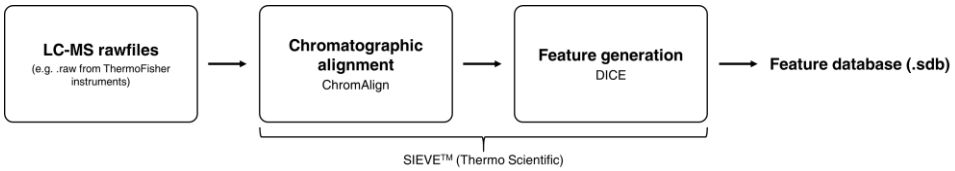
As shown in the figure above, LC-MS raw files are processed in two separated routes to generate protein identification and quantification results. In one route, rawfiles are imported into database search engines for peptide identification, and generated search results are processed for FDR control, peptide inference, and protein group. The final output is a ***.csv spectrum report*** containing the filtered PSM list. In the other route, SIEVE™ is used for quantitative feature generation, including ChromAlign-based RT correction and direct ion-current extraction (DICE). The features are stored in an ***.sdb SQL database.*** The spectrum report and feature database are merged by a customized R package, IonStarStat, to match features with corresponding PSMs. The IonStarStat packages also serves to perform dataset-wide normalization, multivariate algorithm-based outlier rejection, and data aggregation.

FUNCTIONAL MODULES

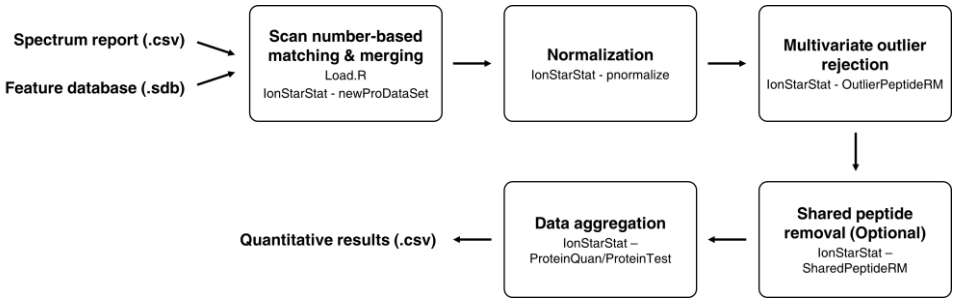
Module 1: Protein Identification



Module 2: Quantitative feature extraction



Module 3: Data integration and quantification



Module 4: Post-quantification data processing

Processing of quantitative results can be achieved by StarGazer, an R-Shiny App currently under development. StarGazer will be available in the next build of IonStar.

SOFTWARE AVAILABILITY

SIEVE™ v2.2 SP2 is a commercial software from Thermo Fisher Scientific. Please contact Thermo Fisher Scientific regarding the quote for SIEVE™. To ensure of proper performance of SIEVE™, we recommend running SIEVE™ on a PC with at least 16-core processors and at least 192 GB RAM.

R package IonStarStat and related scripts (IonStarSPG.R, Load.R, Run IonStar code.R) can be downloaded from https://github.com/shichens1989/IonStarStat. All operations in this manual are accomplished under R version 3.2.4 and Rstudio ver1.0.136.

INSTALLATION INSTRUCTIONS

IonStarStat package can be installed directly in RStuidio by running the following command in the R Console:

```
#####Install dependency and ion star(the beta version called MStest)#####
#Install dependencies "RSQLite""MCMCglmm""affyPLM""mvoutliers"
source("https://bioconductor.org/biocLite.R")
biocLite("affyPLM")
biocLite("MCMCglmm")
biocLite("RSQlite")
biocLite("mvoutliers")
Install.packages("File Directory\\IonStarStat_0.1.4.tar.gz")
```

Alternatively, the **Install Packages** under the **Tools** menu can be used for installation. Yet the supporting packages need to be installed manually as well.

MODULE 1: PROTEIN IDENTIFICATION

The first input into the IonStarStat package is a spectrum report containing a filtered PSM list. This report can be exported conveniently from a number of softwares, e.g. Proteome Discoverer, Scaffold. Here we will demonstrate the use of the IonStarSPG.R spectrum report generator using results obtained from MS-GF+ (database search) and IDPicker (post-search processing).

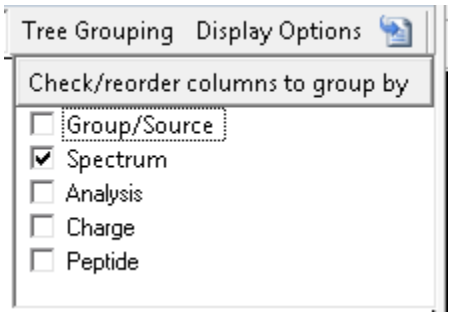
I. Input files

Search results .tsv files **(Sample name.raw.tsv)** converted from .mzid files using MSGFPlus.jar (-showQValue = 1, -shoeDecoy = 1, ⁻unroll = 1)

Protein list .tsv file exported from the Protein View tab of the .idpDB file

Peptide list .tsv file exported from the Peptide View tab of the .idpDB file

Spectrum list .tsv file exported from the Spectrum View tab of the .idpDB file (Tree grouping = Spectrum)



II. Spectrum report generation with IonStarSPG.R

1. Confirm that all input files are located under the same directory.

2. Generate spectrum report

a. Change the input and output file names:

Line 4~7

```
4  setwd("File Directory")
5  protein <- read.csv("Protein list.tsv",header=TRUE,sep="\t",fileEncoding="windows-1252")
6  peptide <- read.table("Peptide list.tsv",header=TRUE,sep="\t",fileEncoding="windows-1252")
7  spectrum <- read.table("Spectra list.tsv",header=TRUE,sep="\t",fileEncoding="windows-1252")
```

Line 70

```
70  write.csv(spectrumrepo_total, "Spectrum report.csv", row.names=FALSE)
```

b. Run the script by **Source** function.

c. The output (Spectrum report.csv) contains the filtered PSM from each LC-MS run as follows:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
1	SpecFile	SpecID	ScanNum	FragMeth	Precursor	IsotopeEr	PrecursorI	Charge	DeNovoSc	MSGFScor	SpecEVal	EValue	QValue	PepQValu	Peptide.S	Accession	Cluster	Count	Coverage	Protein.Gi	Distinct.P	Distinct.M	Filtered.S	Descriptic	Accession.2		
2	171103_IA	scan=2719	27198	HCD	612.5788	1	2.617393	4	166	163	9.65E-29	1.80E-21	0	0	IVVVWVG	sp|Q6120	1	1	58.18966	1	8	8	136	Platelet-a	sp|Q61205|PA1B3_MOUSE		
3	171103_IA	scan=1134	11340	HCD	1134.528	1	-3.84731	2	165	162	1.34E-25	2.51E-18	0	0	SGEGENP	sp|Q6120	1	1	58.18966	1	8	8	136	Platelet-a	sp|Q61205|PA1B3_MOUSE		
4	171103_IA	scan=5277	52776	HCD	483.329	0	0.126281	2	111	111	3.60E-11	0.000656	0.000104	0.00015	VVVLGLL	sp|Q6120	1	1	58.18966	1	8	8	136	Platelet-a	sp|Q61205|PA1B3_MOUSE		
5	171103_IA	scan=1136	11364	HCD	1134.029	0	-1.07643	2	183	175	8.59E-25	1.60E-17	0	0	SGEGENP	sp|Q6120	1	1	58.18966	1	8	8	136	Platelet-a	sp|Q61205|PA1B3_MOUSE		
6	171103_IA	scan=4358	43586	HCD	867.5212	1	-0.10438	2	138	99	2.79E-12	5.19E-05	0	0	LPLHTLTS	sp|Q8BK0	2	1	16.31579	2	2	2	16	Transmem	sp|Q8BK08|TMM11_MOUSI		
7	171103_IA	scan=5969	59690	HCD	833.419	0	-12.9624	2	109	61	4.40E-12	8.14E-05	0	0	EFQEGPLL	sp|Q8C5L	3	1	10.47009	3	3	3	58	Inositol p	sp|Q8C5L6|INP5K_MOUSE		
8	171103_IA	scan=5693	56931	HCD	861.4202	0	-1.06281	2	149	125	2.04E-15	3.79E-08	0	0	ICQSMGLT	sp|O0879	4	1	4.41989	4	2	2	15	Eukaryoti	sp|O08796|EF2K_MOUSE		
9	171103_IA	scan=1284	12843	HCD	616.3039	0	0.396137	2	115	109	5.20E-14	9.55E-07	0	0	AGQFTLGF	sp|P11672	6	1	25	6	5	5	132	Neutroph	sp|P11672|NGAL_MOUSE		
10	171103_IA	scan=4309	43094	HCD	702.8854	0	0	2	139	134	3.41E-18	6.30E-11	0	0	WYVVGLA	sp|P11672	6	1	25	6	5	5	132	Neutroph	sp|P11672|NGAL_MOUSE		
11	171103_IA	scan=2348	23484	HCD	431.2393	-1	-4.03146	2	69	69	1.12E-10	0.002022	0.000277	0.000401	IADFGLAR	sp|P39686	8	1	21.04283	8	9	9	141	Tyrosine-p	sp|P39688|FYN_MOUSE		
12	171103_IA	scan=7987	7987	HCD	612.3014	0	4.286326	2	115	115	2.09E-13	3.83E-06	0	0	LIEDNEYT	sp|P39686	8	1	21.04283	8	9	9	141	Tyrosine-p	sp|P39688|FYN_MOUSE		
13	171103_IA	scan=3083	30834	HCD	618.3144	1	-2.71508	2	122	120	2.84E-15	5.22E-08	0	0	WTAPEAA	sp|P39686	8	1	21.04283	8	9	9	141	Tyrosine-p	sp|P39688|FYN_MOUSE		
14	171103_IA	scan=2488	24883	HCD	431.7436	0	2.120539	2	103	103	2.23E-10	0.004044	0.000592	0.000401	IADFGLAR	sp|P39686	8	1	21.04283	8	9	9	141	Tyrosine-p	sp|P39688|FYN_MOUSE		
15	171103_IA	scan=2383	23837	HCD	774.0774	0	1.103886	3	179	105	9.05E-14	1.69E-06	0	0	MERPGPG	sp|Q9JHU	9	1	42.90844	9	17	17	549	Inositol-3	sp|Q9JHU9|INO1_MOUSE		
16	171103_IA	scan=3101	31011	HCD	1054.183	0	1.042167	3	243	93	1.40E-13	2.63E-06	0	0	SSVVDDM	sp|Q9JHU	9	1	42.90844	9	17	17	549	Inositol-3	sp|Q9JHU9|INO1_MOUSE		
17	171103_IA	scan=5282	52828	HCD	888.4451	0	0.961785	3	244	107	5.79E-13	1.08E-05	0	0	AQVLDCG	sp|Q9JHU	9	1	42.90844	9	17	17	549	Inositol-3	sp|Q9JHU9|INO1_MOUSE		
18	171103_IA	scan=3096	30962	HCD	632.9118	0	-0.09644	5	215	162	2.46E-22	4.61E-15	0	0	SSVVDDM	sp|Q9JHU	9	1	42.90844	9	17	17	549	Inositol-3	sp|Q9JHU9|INO1_MOUSE		
19	171103_IA	scan=5066	50669	HCD	1388.681	0	2.461313	2	189	162	8.34E-22	1.56E-14	0	0	TMSIVSYN	sp|Q9JHU	9	1	42.90844	9	17	17	549	Inositol-3	sp|Q9JHU9|INO1_MOUSE		
20	171103_IA	scan=8006	80069	HCD	674.3995	0	2.53409	2	131	126	7.32E-15	1.35E-07	0	0	SVLVDFLI	sp|Q9JHU	9	1	42.90844	9	17	17	549	Inositol-3	sp|Q9JHU9|INO1_MOUSE		
21	171103_IA	scan=5283	52834	HCD	666.5853	0	0.549304	4	235	206	4.48E-22	8.36E-15	0	0	AQVLDCG	sp|Q9JHU	9	1	42.90844	9	17	17	549	Inositol-3	sp|Q9JHU9|INO1_MOUSE		
22	171103_IA	scan=5302	53028	HCD	666.3282	-1	-8.90538	4	200	130	2.17E-17	4.04E-10	0	0	AQVLDCG	sp|Q9JHU	9	1	42.90844	9	17	17	549	Inositol-3	sp|Q9JHU9|INO1_MOUSE		
23	171103_IA	scan=4952	49525	HCD	853.4715	0	0.786655	2	188	183	6.51E-19	1.21E-11	0	0	NLNQALL	sp|P29391	10	1	65.02732	10	13	13	547	Ferritin lig	sp|P29391|FRIL1_MOUSE		
24	171103_IA	scan=6369	63698	HCD	820.899	1	1.506723	4	264	261	2.12E-36	3.97E-29	0	0	ALFQDVQ	sp|P29391	10	1	65.02732	10	13	13	547	Ferritin lig	sp|P29391|FRIL1_MOUSE		
25	171103_IA	scan=4946	49467	HCD	569.3164	0	0.214416	3	134	130	2.82E-19	5.24E-12	0	0	NLNQALL	sp|P29391	10	1	65.02732	10	13	13	547	Ferritin lig	sp|P29391|FRIL1_MOUSE		
26	171103_IA	scan=2634	26348	HCD	938.467	0	1.690967	2	156	150	1.77E-19	3.28E-12	0	0	ALFQDVQ	sp|P29391	10	1	65.02732	10	13	13	547	Ferritin lig	sp|P29391|FRIL1_MOUSE		
27	171103_IA	scan=5168	51685	HCD	791.7334	0	1.002178	3	179	165	4.42E-22	8.25E-15	0	0	VAGPQPA	sp|P29391	10	1	65.02732	10	13	13	547	Ferritin lig	sp|P29391|FRIL1_MOUSE		
28	171103_IA	scan=2887	28874	HCD	711.8348	0	1.45764	2	135	132	4.00E-17	7.38E-10	0	0	TQEAMEA	sp|P29391	10	1	65.02732	10	13	13	547	Ferritin lig	sp|P29391|FRIL1_MOUSE		
29	171103_IA	scan=5162	51628	HCD	1187.596	1	-0.59039	2	250	236	2.68E-24	5.01E-17	0	0	VAGPQPA	sp|P29391	10	1	65.02732	10	13	13	547	Ferritin lig	sp|P29391|FRIL1_MOUSE		
30	171103_IA	scan=6375	63755	HCD	1094.193	1	-1.6547	3	255	212	1.70E-25	3.19E-18	0	0	ALFQDVQ	sp|P29391	10	1	65.02732	10	13	13	547	Ferritin lig	sp|P29391|FRIL1_MOUSE		
31	171103_IA	scan=6314	6314	HCD	528.2744	0	-0.11554	2	98	97	5.05E-13	9.21E-06	0	0	MGNHLTN	sp|P29391	10	1	65.02732	10	13	13	547	Ferritin lig	sp|P29391|FRIL1_MOUSE		
32	171103_IA	scan=1921	19214	HCD	740.8408	0	-19.1133	2	117	75	1.72E-10	0.003182	0.000422	0.000114	QNYSTEVE	sp|P29391	10	1	65.02732	10	13	13	547	Ferritin lig	sp|P29391|FRIL1_MOUSE		
33	171103_IA	scan=1856	18569	HCD	494.2394	0	0.740959	3	106	76	2.91E-11	0.000537	7.90E-05	0.000114	QNYSTEVE	sp|P29391	10	1	65.02732	10	13	13	547	Ferritin lig	sp|P29391|FRIL1_MOUSE		
34	171103_IA	scan=5217	52172	HCD	1187.596	1	-0.59039	2	250	243	6.38E-25	1.19E-17	0	0	VAGPQPA	sp|P29391	10	1	65.02732	10	13	13	547	Ferritin lig	sp|P29391|FRIL1_MOUSE		
35	171103_IA	scan=4947	49476	HCD	569.3166	0	0.536039	3	167	161	6.39E-19	1.19E-11	0	0	NLNQALL	sp|P29391	10	1	65.02732	10	13	13	547	Ferritin lig	sp|P29391|FRIL1_MOUSE		
36	171103_IA	scan=6363	63637	HCD	1093.855	0	-4.91022	3	128	33	5.09E-12	9.53E-05	0	0	ALFQDVQ	sp|P29391	10	1	65.02732	10	13	13	547	Ferritin lig	sp|P29391|FRIL1_MOUSE		
37	171103_IA	scan=2634	26348	HCD	938.467	0	1.690967	2	156	150	1.77E-19	3.28E-12	0	0	ALFQDVQ	sp|P49945	11	1	30.60109	11	7	7	266	Ferritin lig	sp|P49945|FRIL2_MOUSE		
38	171103_IA	scan=1856	18569	HCD	494.2394	0	0.740959	3	106	76	2.91E-11	0.000537	7.90E-05	0.000114	QNYSTEVE	sp|P49945	10	1	30.60109	11	7	7	266	Ferritin lig	sp|P49945|FRIL2_MOUSE		
39	171103_IA	scan=4946	49467	HCD	569.3164	0	0.214416	3	134	130	2.82E-19	5.24E-12	0	0	NLNQALL	sp|P49945	10	1	30.60109	11	7	7	266	Ferritin lig	sp|P49945|FRIL2_MOUSE		
40	171103_IA	scan=4947	49476	HCD	569.3166	0	0.536039	3	167	161	6.39E-19	1.19E-11	0	0	NLNQALL	sp|P49945	10	1	30.60109	11	7	7	266	Ferritin lig	sp|P49945|FRIL2_MOUSE		
41	171103_IA	scan=4952	49525	HCD	853.4715	0	0.786655	2	188	183	6.51E-19	1.21E-11	0	0	NLNQALL	sp|P49945	10	1	30.60109	11	7	7	266	Ferritin lig	sp|P49945|FRIL2_MOUSE		
42	171103_IA	scan=6314	6314	HCD	528.2744	0	-0.11554	2	98	97	5.05E-13	9.21E-06	0	0	MGNHLTN	sp|P49945	10	1	30.60109	11	7	7	266	Ferritin lig	sp|P49945|FRIL2_MOUSE		
43	171103_IA	scan=1921	19214	HCD	740.8408	0	-19.1133	2	117	75	1.72E-10	0.003182	0.000422	0.000114	QNYSTEVE	sp|P49945	10	1	30.60109	11	7	7	266	Ferritin lig	sp|P49945|FRIL2_MOUSE		
44	171103_IA	scan=2381	23812	HCD	687.3893	0	2.752582	2	148	131	4.41E-13	8.13E-06	0	0	IEVEKPFA	sp|P24369	11	1	63.88889	12	20	20	979	Peptidyl-p	sp|P24369|PPIB_MOUSE		
45	171103_IA	scan=4638	46383	HCD	729.8593	1	-13.0948	2	155	130	3.55E-14	6.56E-07	0	0	DTNGSQFI	sp|P24369	11	1	63.88889	12	20	20	979	Peptidyl-p	sp|P24369|PPIB_MOUSE		

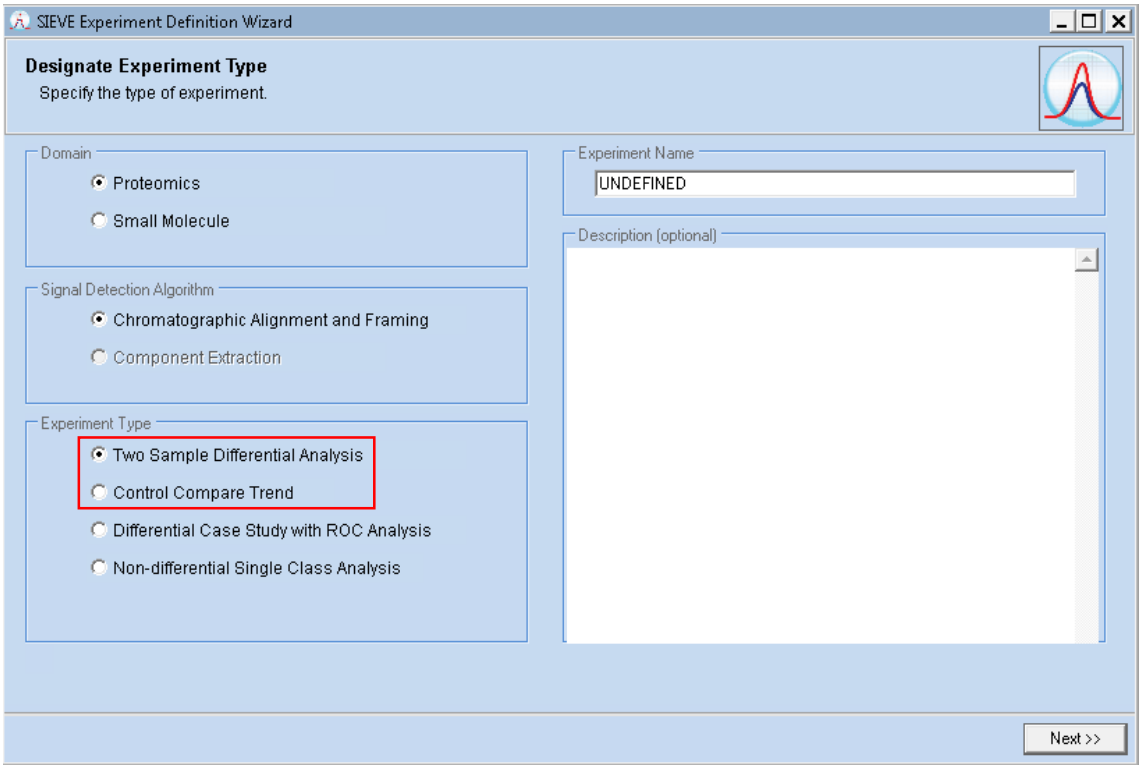
MODULE 2: QUANTITATIVE FEATURE GENERATION

Quantitative feature generation in IonStar is accomplished by SIEVE™ v2.2 SP2 (Thermo Scientific), which integrates ChromAlign for global 3-D chromatographic alignment and a DICE method for feature extraction.

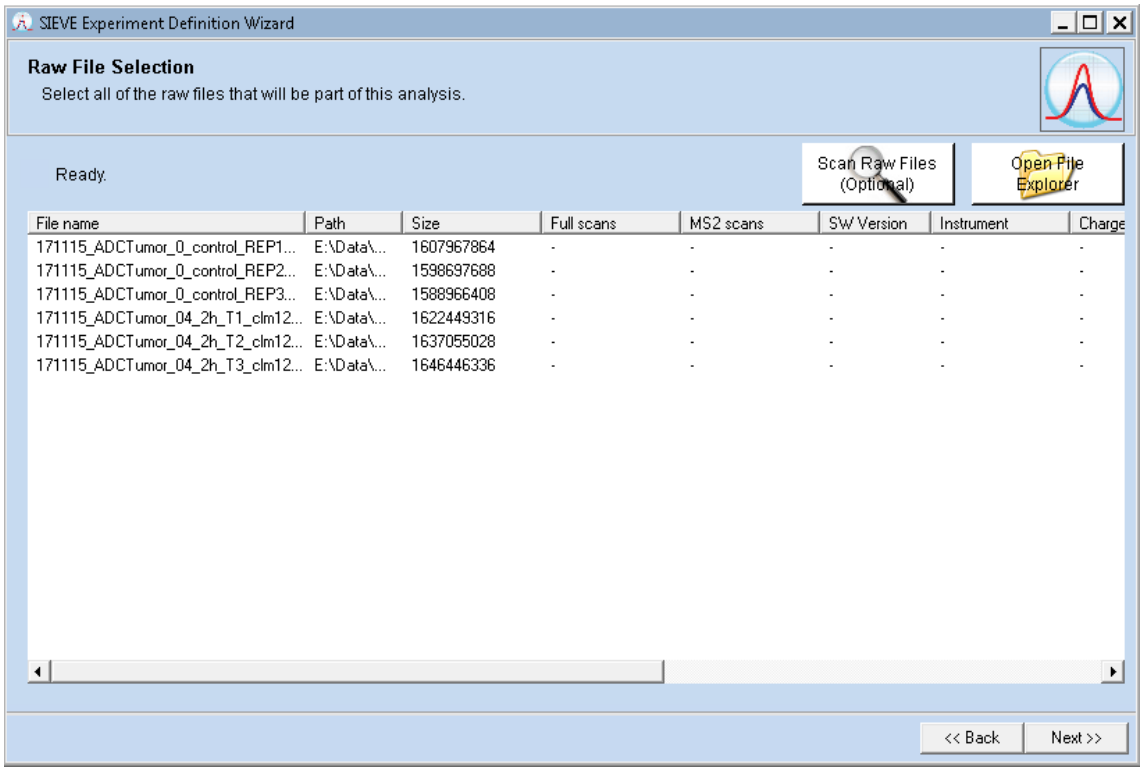
I. Setting up the method

1. Load rawfiles into SIEVE™

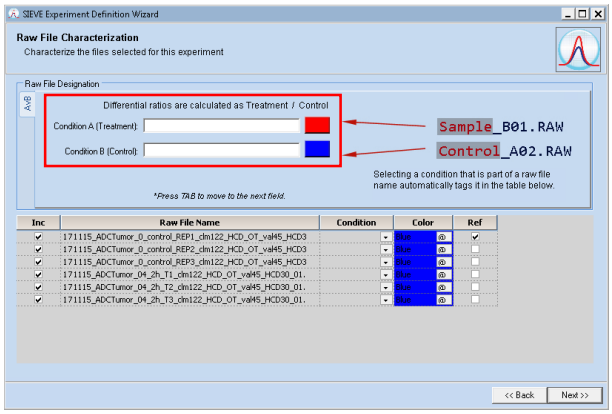
To start the quantitative feature generation analysis, open SIEVE™ and select **File → Create new experiment**. On the **Designate Experiment Type** page, select the Experiment Type based on the study. For a case-control experiment, use **Two Sample Differential Analysis**; for multi-condition experiment (≥3 conditions including control), use **Control Compare Trend**.

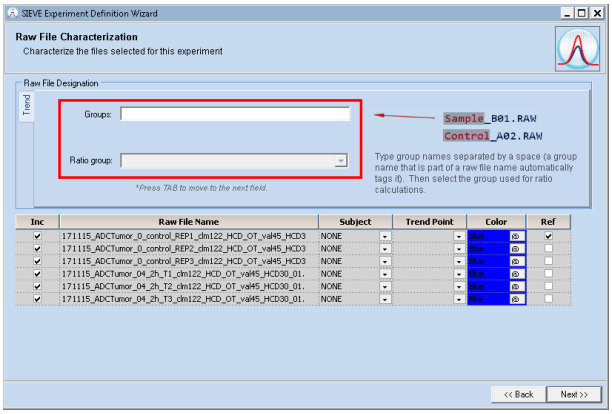


Drag all rawfiles onto the **Raw File Selection** page.

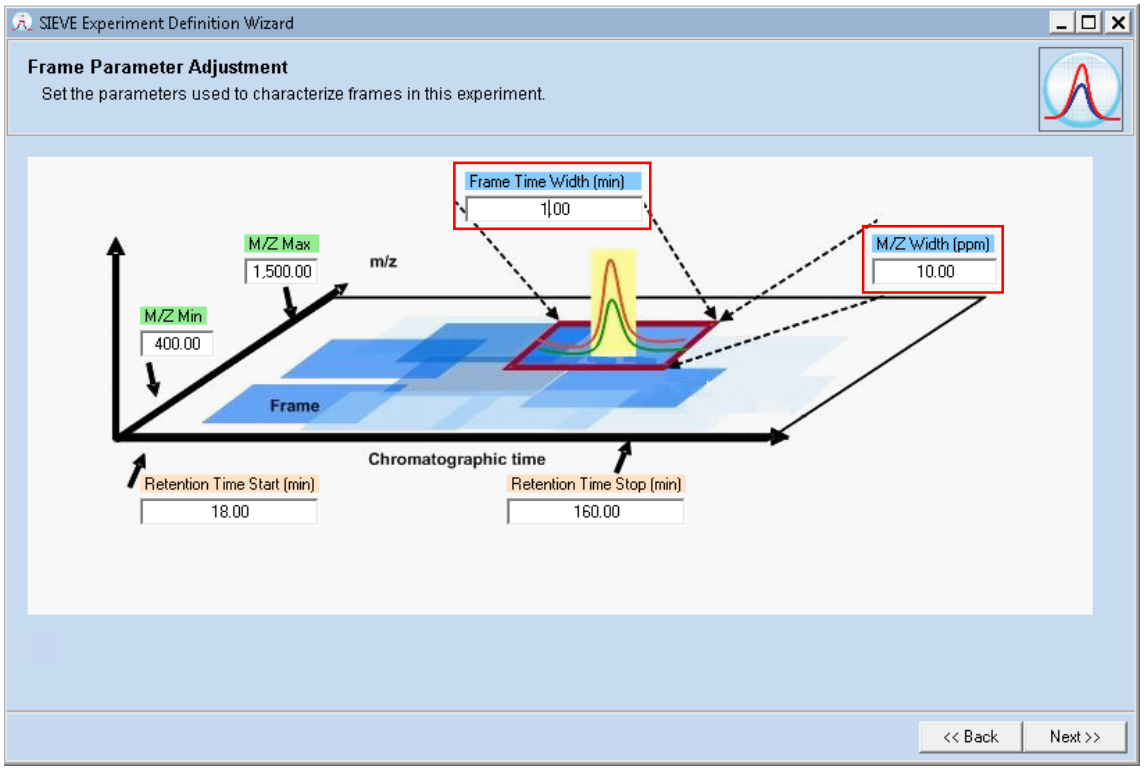


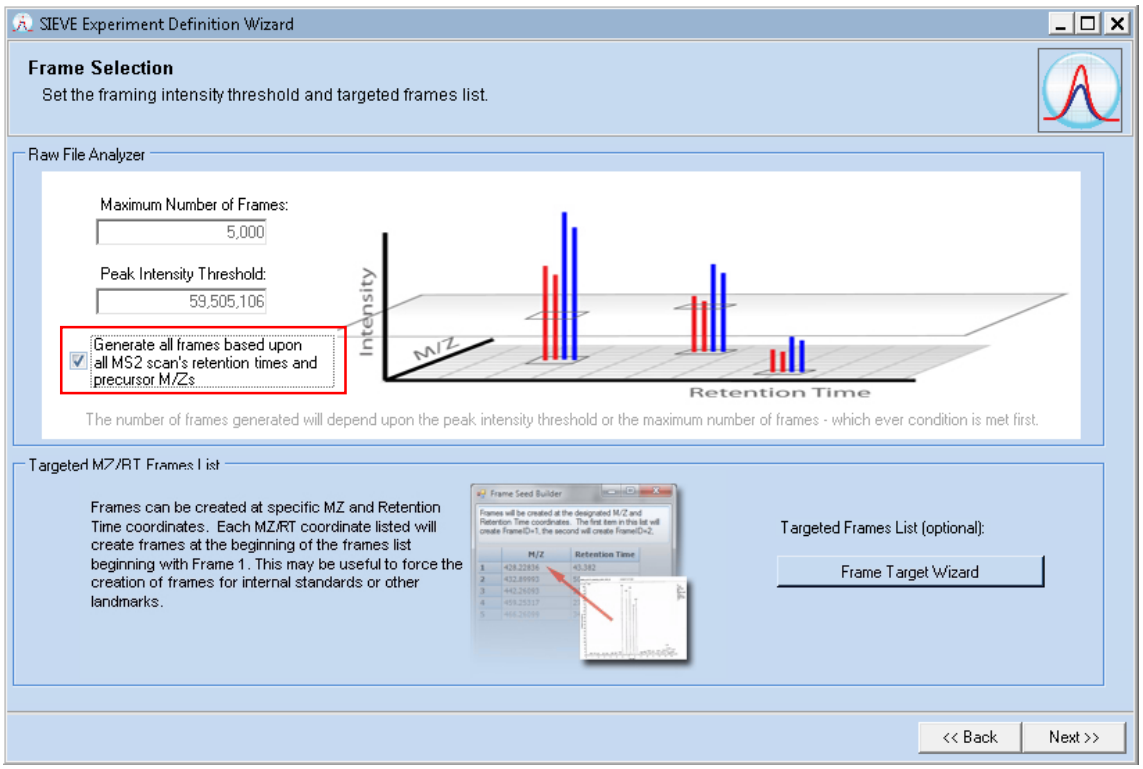
Assign conditions to each file imported. For **Two Sample Differential Analysis**, assign Condition A and B in the two boxes; for **Control Compare Trend**, assign all conditions in the upper box and the control condition in the lower box.



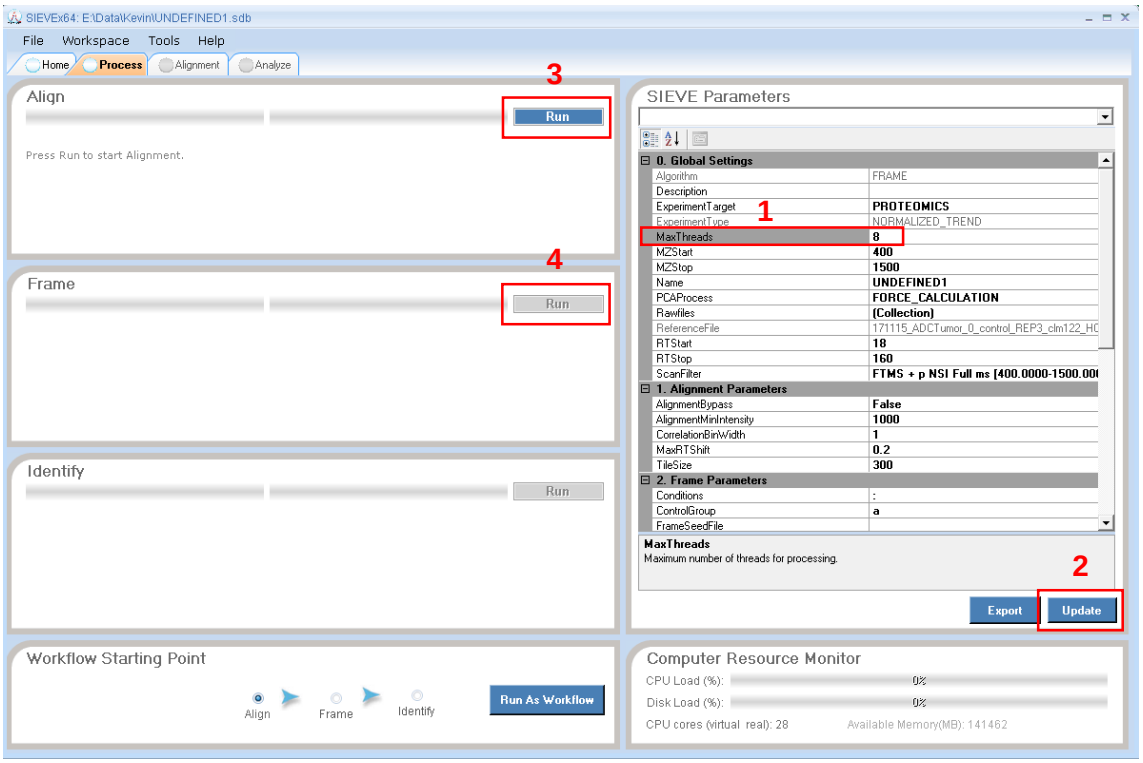


Setup the method parameters. The parameters that needs to be modified (for LC-MS rawfiles acquired on a Thermo Orbitrap instrument under 120K MS1 resolution) are shown as below. All other parameters follow the default settings.





After loading files with the given parameters, For IonStar, users do not need to run the Identify process. MaxThreads also needs to be modified according to the configuration of the computer used for SIEVE™. For example, 6~8 threads are recommended for a computer with 16-core processors and 92 GB RAM. A suggested sequence on the **Process** tab is highlighted in the following figure.



After the Frame process, an .sdb file will be generated containing all quantitative features. For more detailed information about the use of SIEVE™, please refer to SIEVE™ User Guide:

https://tools.thermofisher.com/content/sfs/manuals/Man-XCALI-97696-SIEVE-22-User-ManXCALI97696-A-EN.pdf

MODULE 3: DATA INTEGRATION AND QUANTIFICATION

After protein identification and quantitative feature generation, the customized R package IonStarStat plus associating R scripts will be utilized to generate the final quantitative results. Steps involved in this module include:

- Generation of the annotated frame list by integrating the spectrum report and the SIEVE database;
- Removal of redundant frames;
- Aggregation of frame-level data to peptide level & inter-sample normalization of quantitative intensities;
- Multivariate mean variation-based outlier peptide detection;
- Shared peptide removal (optional);
- Peptide-to-protein aggregation via two different algorithms.

The instructions for each step using IonStarStat will be elaborated as follows.

I. Generation of the annotated frame list

1. Required packages for frame list annotation

To generate the frame list, the following R packages need to be installed: **XLConnect**, **RSQLite**.

2. Specify the input and output files

Input files: Line 1~5

```
1  setwd("File directory")
2  ##File name of SIEVE file
3  db<-"SIEVE database.sdb"
4  ##File name of Spectrum report
5  xls<-"Spectrum report.csv"
```

Output files: Line 37~38 or Line 71~72

```
37  write.csv(sub_data,"Annotated Frame List.csv",row.names=FALSE)
38  write.csv(colnames(sub_data),"Sample List.csv")
```

3. Generate the annotated frame list & the sample list

Generate the annotated frame list using the script.

Line 7~37 summarize scripts used for spectrum reports exported from Scaffold.

Line 39~70 summarize scripts used for spectrum reports merged by IonStarSPG.R.

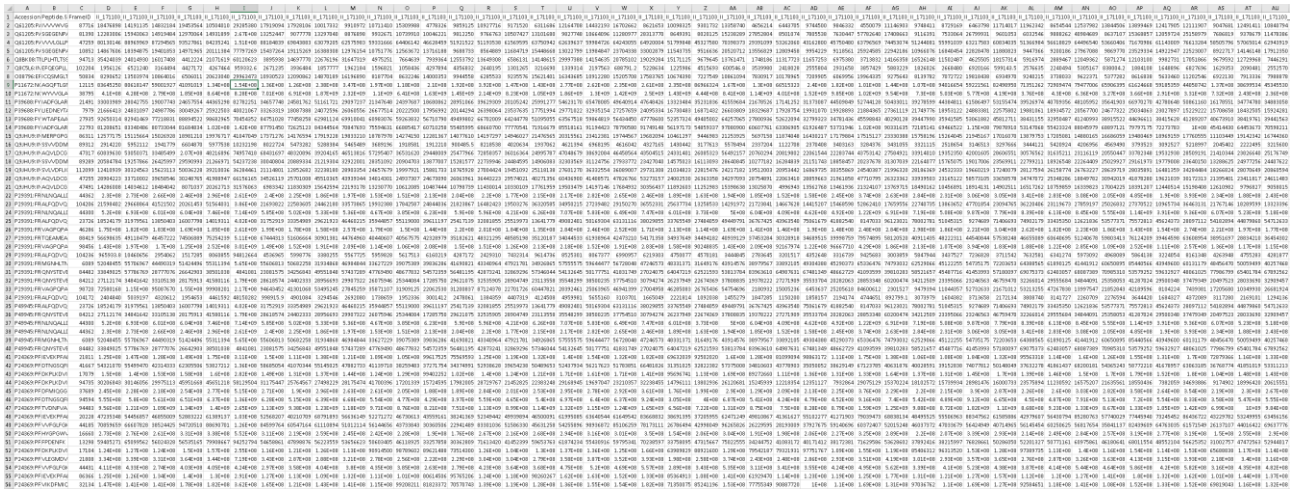
Note that Line 36 or 70 needs to be modified.

```
36  sub_data<-FS[,"columns"]
```

The columns to be included are **Peptide AC, Peptide sequence, Frame ID, and Frame intensity in each sample**.

- If other software combinations are used to generate the spectrum report, the scripts can be modified so that column headers are consistent between the spectrum report and SIEVE database for matching (recommended for users familiar with R language).

An example of the exported Frame list is shown as follows:



An example of the exported Sample list is shown as follows:

	A	B	C	D	E	F	G	H
1		x						
2		1 Accession						
3		2 Peptide.Sequence						
4		3 FrameID						
5		4 II_171103_IAVIung_00_REP4_clm122_HCD_OT_val45_HCD30_01_4ul						
6		5 II_171103_IAVIung_00A_clm122_HCD_OT_val45_HCD30_01_4ul						
7		6 II_171103_IAVIung_00B_clm122_HCD_OT_val45_HCD30_01_4ul						
8		7 II_171103_IAVIung_00C_clm122_HCD_OT_val45_HCD30_01_4ul						
9		8 II_171103_IAVIung_01A_clm122_HCD_OT_val45_HCD30_01_4ul						
10		9 II_171103_IAVIung_01B_clm122_HCD_OT_val45_HCD30_01_4ul						
11		10 II_171103_IAVIung_01C_clm122_HCD_OT_val45_HCD30_01_4ul						
12		11 II_171103_IAVIung_01D_clm122_HCD_OT_val45_HCD30_01_4ul						
13		12 II_171103_IAVIung_02A_clm122_HCD_OT_val45_HCD30_01_4ul						
14		13 II_171103_IAVIung_02B_clm122_HCD_OT_val45_HCD30_01_4ul						
15		14 II_171103_IAVIung_02C_clm122_HCD_OT_val45_HCD30_01_4ul						
16		15 II_171103_IAVIung_02D_clm122_HCD_OT_val45_HCD30_01_4ul						
17		16 II_171103_IAVIung_03A_clm122_HCD_OT_val45_HCD30_01_4ul						
18		17 II_171103_IAVIung_03B_clm122_HCD_OT_val45_HCD30_01_4ul						
19		18 II_171103_IAVIung_03C_clm122_HCD_OT_val45_HCD30_01_4ul						
20		19 II_171103_IAVIung_03D_clm122_HCD_OT_val45_HCD30_01_4ul						
21		20 II_171103_IAVIung_04A_clm122_HCD_OT_val45_HCD30_01_4ul						
22		21 II_171103_IAVIung_04B_clm122_HCD_OT_val45_HCD30_01_4ul						
23		22 II_171103_IAVIung_04C_clm122_HCD_OT_val45_HCD30_01_4ul						
24		23 II_171103_IAVIung_04D_clm122_HCD_OT_val45_HCD30_01_4ul						
25		24 II_171103_IAVIung_05A_clm122_HCD_OT_val45_HCD30_01_4ul						
26		25 II_171103_IAVIung_05B_clm122_HCD_OT_val45_HCD30_01_4ul						
27		26 II_171103_IAVIung_05C_clm122_HCD_OT_val45_HCD30_01_4ul						
28		27 II_171103_IAVIung_05D_clm122_HCD_OT_val45_HCD30_01_4ul						
29		28 II_171103_IAVIung_06A_clm122_HCD_OT_val45_HCD30_01_4ul						
30		29 II_171103_IAVIung_06B_clm122_HCD_OT_val45_HCD30_01_4ul						
31		30 II_171103_IAVIung_06C_clm122_HCD_OT_val45_HCD30_01_4ul						
32		31 II_171103_IAVIung_06D_clm122_HCD_OT_val45_HCD30_01_4ul						
33		32 II_171103_IAVIung_07A_clm122_HCD_OT_val45_HCD30_01_4ul						
34		33 II_171103_IAVIung_07B_clm122_HCD_OT_val45_HCD30_01_4ul						
35		34 II_171103_IAVIung_07C_clm122_HCD_OT_val45_HCD30_01_4ul						
36		35 II_171103_IAVIung_07D_clm122_HCD_OT_val45_HCD30_01_4ul						
37		36 II_171103_IAVIung_08A_clm122_HCD_OT_val45_HCD30_01_4ul						
38		37 II_171103_IAVIung_08B_clm122_HCD_OT_val45_HCD30_01_4ul						
39		38 II_171103_IAVIung_08C_clm122_HCD_OT_val45_HCD30_01_4ul						
40		39 II_171103_IAVIung_08D_clm122_HCD_OT_val45_HCD30_01_4ul						
41		40 II_171103_IAVIung_09A_clm122_HCD_OT_val45_HCD30_01_4ul						
42		41 II_171103_IAVIung_09B_clm122_HCD_OT_val45_HCD30_01_4ul						
43		42 II_171103_IAVIung_09C_clm122_HCD_OT_val45_HCD30_01_4ul						
44		43 II_171103_IAVIung_09D_clm122_HCD_OT_val45_HCD30_01_4ul						
45		44 II_171103_IAVIung_10A_clm122_HCD_OT_val45_HCD30_01_4ul						
46		45 II_171103_IAVIung_10B_clm122_HCD_OT_val45_HCD30_01_4ul						
47		46 II_171103_IAVIung_10C_clm122_HCD_OT_val45_HCD30_01_4ul						
48		47 II_171103_IAVIung_10D_clm122_HCD_OT_val45_HCD30_01_4ul						

II. Protein quantification

1. Input files & load IonStarStat

Annotated frame list (.csv) A list of filtered frames with annotation (peptide sequence and protein AC).

Grouped sample list (.txt) A list containing all sample names and corresponding group information.

To generate the group list for quantification, open the sample list file (.csv) in excel. The file should look like this:

	A	B
1	Rawfilles	GroupID
2	II_171103_IAVlung_00_REP4_clm122_HCD_OT_val45_HCD30_01_4ul	A
3	II_171103_IAVlung_00A_clm122_HCD_OT_val45_HCD30_01_4ul	A
4	II_171103_IAVlung_00B_clm122_HCD_OT_val45_HCD30_01_4ul	A
5	II_171103_IAVlung_00C_clm122_HCD_OT_val45_HCD30_01_4ul	A
6	II_171103_IAVlung_01A_clm122_HCD_OT_val45_HCD30_01_4ul	B
7	II_171103_IAVlung_01B_clm122_HCD_OT_val45_HCD30_01_4ul	B
8	II_171103_IAVlung_01C_clm122_HCD_OT_val45_HCD30_01_4ul	B
9	II_171103_IAVlung_01D_clm122_HCD_OT_val45_HCD30_01_4ul	B
10	II_171103_IAVlung_02A_clm122_HCD_OT_val45_HCD30_01_4ul	C
11	II_171103_IAVlung_02B_clm122_HCD_OT_val45_HCD30_01_4ul	C
12	II_171103_IAVlung_02C_clm122_HCD_OT_val45_HCD30_01_4ul	C
13	II_171103_IAVlung_02D_clm122_HCD_OT_val45_HCD30_01_4ul	C
14	II_171103_IAVlung_03A_clm122_HCD_OT_val45_HCD30_01_4ul	D
15	II_171103_IAVlung_03B_clm122_HCD_OT_val45_HCD30_01_4ul	D
16	II_171103_IAVlung_03C_clm122_HCD_OT_val45_HCD30_01_4ul	D
17	II_171103_IAVlung_03D_clm122_HCD_OT_val45_HCD30_01_4ul	D
18	II_171103_IAVlung_04A_clm122_HCD_OT_val45_HCD30_01_4ul	E
19	II_171103_IAVlung_04B_clm122_HCD_OT_val45_HCD30_01_4ul	E
20	II_171103_IAVlung_04C_clm122_HCD_OT_val45_HCD30_01_4ul	E
21	II_171103_IAVlung_04D_clm122_HCD_OT_val45_HCD30_01_4ul	E
22	II_171103_IAVlung_05A_clm122_HCD_OT_val45_HCD30_01_4ul	F
23	II_171103_IAVlung_05B_clm122_HCD_OT_val45_HCD30_01_4ul	F
24	II_171103_IAVlung_05C_clm122_HCD_OT_val45_HCD30_01_4ul	F
25	II_171103_IAVlung_05D_clm122_HCD_OT_val45_HCD30_01_4ul	F
26	II_171103_IAVlung_06A_clm122_HCD_OT_val45_HCD30_01_4ul	G
27	II_171103_IAVlung_06B_clm122_HCD_OT_val45_HCD30_01_4ul	G
28	II_171103_IAVlung_06C_clm122_HCD_OT_val45_HCD30_01_4ul	G
29	II_171103_IAVlung_06D_clm122_HCD_OT_val45_HCD30_01_4ul	G
30	II_171103_IAVlung_07A_clm122_HCD_OT_val45_HCD30_01_4ul	H
31	II_171103_IAVlung_07B_clm122_HCD_OT_val45_HCD30_01_4ul	H
32	II_171103_IAVlung_07C_clm122_HCD_OT_val45_HCD30_01_4ul	H
33	II_171103_IAVlung_07D_clm122_HCD_OT_val45_HCD30_01_4ul	H
34	II_171103_IAVlung_08A_clm122_HCD_OT_val45_HCD30_01_4ul	I
35	II_171103_IAVlung_08B_clm122_HCD_OT_val45_HCD30_01_4ul	I
36	II_171103_IAVlung_08C_clm122_HCD_OT_val45_HCD30_01_4ul	I
37	II_171103_IAVlung_08D_clm122_HCD_OT_val45_HCD30_01_4ul	I
38	II_171103_IAVlung_09A_clm122_HCD_OT_val45_HCD30_01_4ul	J
39	II_171103_IAVlung_09B_clm122_HCD_OT_val45_HCD30_01_4ul	J
40	II_171103_IAVlung_09C_clm122_HCD_OT_val45_HCD30_01_4ul	J
41	II_171103_IAVlung_09D_clm122_HCD_OT_val45_HCD30_01_4ul	J
42	II_171103_IAVlung_10A_clm122_HCD_OT_val45_HCD30_01_4ul	K
43	II_171103_IAVlung_10B_clm122_HCD_OT_val45_HCD30_01_4ul	K
44	II_171103_IAVlung_10C_clm122_HCD_OT_val45_HCD30_01_4ul	K
45	II_171103_IAVlung_10D_clm122_HCD_OT_val45_HCD30_01_4ul	K

The column headers are **Rawfiles** and **GroupID** (case-sensitive), and samples in the same group should be assigned the same GroupID. The GroupID can be any combinations of alphabets and numbers.

Open **Run IonStar code.R** and run Line 1 to load IonStarStat.

2. Removal of redundant quantitative features (frames)

a. Open **Run IonStar code.R**.

b. Modify the following lines for input and output files:

Line 2~4

```
2  setwd("File directory")
3  rawfile <- "Raw_input_IonStar.csv"
4  condfile <- "Group list.txt"
```

Line 13~14

```
14  wrote.csv(quan,"Protein quantitative result.csv")
15  write.csv(exprs(cdata),"Peptide quantitative result.csv")
```

c. Run Line 5~7 to read the annotated frame list and the grouped sample list into R environment.

It is recommended to check the grouping information by entering **condition** in the R Console:

```
> condition
 [1] A A A A B B B B C C C C D D D D E E E E F F F F G G G G H H H H I I I I J J J J K K K K
Levels: A B C D E F G H I J K
```

d. Run Line 8 to remove redundant frames (i.e. frames assigned to multiple peptide sequences) will also be removed to avoid ambiguity in quantification. The number of protein before and after removal, as well as the number of redundant frames removed will be reported in the console:

```
Input 6527 proteins.
19526 duplicated frames founded.
6465 proteins left after filtering.
```

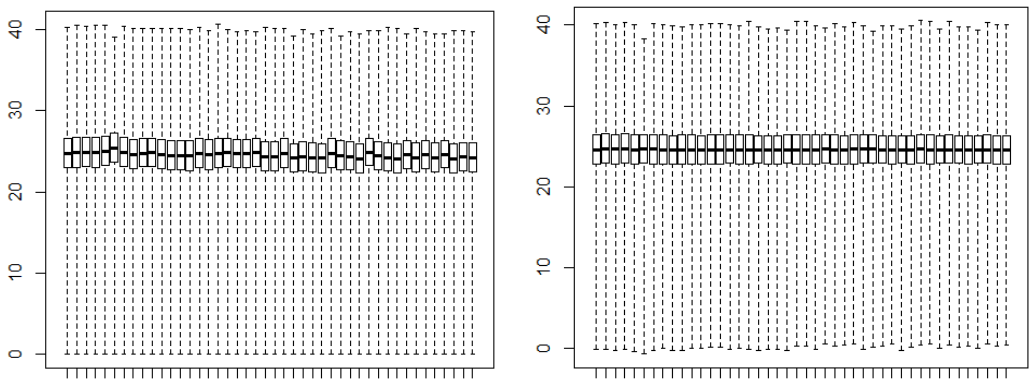
3. Frame-to-peptide aggregation and inter-sample normalization

Run Line 9 to perform frame-to-peptide aggregation and inter-sample normalization of peptide intensities. Parameters include:

summarize = TRUE, FALSE (Whether to aggregate frame-level data to peptide level)

method = NULL, “TIC”, “Quantiles” (Which normalization method is used, TIC being the default)

An example of sample intensities before and after normalization is shown below:



4. Peptide-level outlier detection

Run Line 10 to perform outlier peptide detection. IonStar uses Principal Component-based Outlier Detection (*PCOut*) for outlier detection, which is tailored for multi-condition comparison (≥3 conditions including control).

```
> cdata<-OutlierPeptideRM(ndata,condition,variance=0.7,critM1=1/3,critM2=1/4,ratio=TRUE)
13734 outliers were removed; 46268 peptides left after outlier removal.
```

Parameter **variation** = 0.7~0.9 can be adjusted according to the stringency needed for outlier detection. The higher the value the more outliers will be rejected.

For case-control comparison, please set parameter **ratio** = FALSE. Alternatively, Grubb's test can be used for outlier rejection. The Grubb's test function will be available in the next build of IonStarStat.

NOTE: during this step, errors could occasionally occur if there are too many missing values in certain frames. In this case, run Line 17~20 to remove frames with missing values, then restart the workflow from Line 8.

5. Shared peptide removal (Optional)

Run Line 11 to remove shared peptides (i.e. peptides inferred to ≥2 unique protein groups, a.k.a. degenerate peptides). This step is optional for protein quantification, as many highly abundant proteins share a large proportion of homologous sequence domains. Removal of these peptides could be counterproductive for quantification. However, in specific cases, such as quantification of mixed-species samples, removal of shared peptides with species ambiguity is necessary to obtain species-specific quantitative results.

6. Peptide-to-protein aggregation

Run Line 12 to obtain the quantitative results. Parameters include:

method = “sum”, “fit”

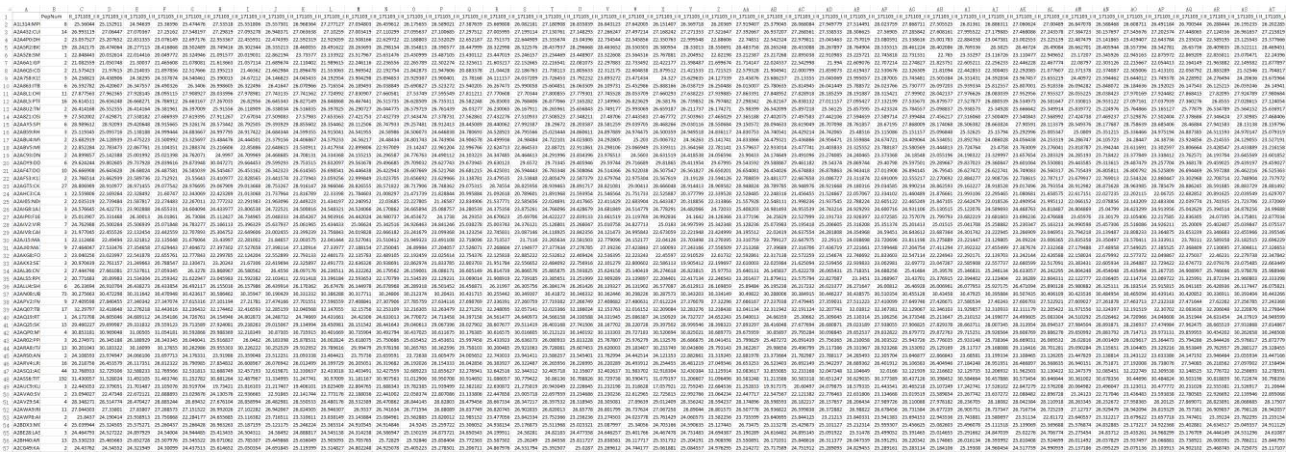
“sum” will simply add the intensities of all peptides inferred to the same protein to obtain protein quantitative intensities.

“fit” will apply General Linear Mixed Model (GLMM) to summarize peptide intensities to protein level.

7. Export quantitative results

Run line 13~14 to export peptide and protein quantitative results (.csv).

The output (Protein quantitative result.csv) contains raw quantitative intensities of all proteins as follows:



MODULE 4: POST-QUANTIFICATION DATA PROCESSING

StarGazer, the module for post-quantification data processing, will be made available in the next build of IonStar.

CONTACT INFORMATION

For questions, suggestions, and other topics about IonStar, feel free to contact us:

Shichen Shen shichens@buffalo.edu

Xue Wang xwang79@buffalo.edu

Jun Qu junqu@buffalo.edu