# Step-by-Step Tutorial for Analyzing DDA and DIA Proteomics Data with Contaminant FASTA and Spectral Libraries

Mass spectrometry-based proteomics is challenged by the presence of contaminant protein background signals. During data analysis, contaminant FASTA libraries allow the search algorithm to distinguish between peptides with similar retention times and m/z. In this study, we generated a custom contaminant FASTA and spectral libraries that can be used for both data-dependent acquisition (DDA) and data-independent acquisition (DIA) software. These new contaminant libraries have been shown to reduce false identifications, increase protein IDs, and do not influence protein quantification for both DIA and DDA workflows. We have also modified the contaminant FASTA library to contain a "Cont" prefix before each UniProt identifier, simplifying the process of removing contaminant proteins prior to statistical analysis.

In this tutorial, we describe how to use our contaminant FASTA library with various DDA and DIA software platforms.

## **Table of Content:**

- 1. Brief Description of Contaminant Libraries
- 2. Removing Protein Contamiannts from Result File in Excel
- 3. Proteome Discoverer for DDA
- 4. MaxQuant for DDA
- 5. MaxDIA for DIA
- 6. Spectronaut for DIA
- 7. DIA-NN for DIA
- 8. Skyline for DIA
- 9. PECAN for DIA

# 1. Brief Description of Contaminant Libraries

Exogenous contaminant proteins orignated from reagents and sample handling are mostly shared in all bottom-up proteomic experiments. Although widely used for DDA proteomics, the list of common protein contaminants from Maxquant and cRAP list have not been updated for years. These contaminant protein lists contain many incorrect Uniprot IDs, some sample-specific interference proteins that are incorrectly listed as contaminants, and available human protein standards from Sigma-Aldrich which are not contaminant proteins. Therefore, we first built a new contaminant FASTA library by manually merging the available contaminant lists online, updating their Uniprot entry IDs, deleting noncontaminant proteins, searching new contaminant proteins on Uniprot, and combining them into a new FASTA file. Our new contaminant FASTA library contains 381 contaminant proteins including all human keratins and skin-derived proteins, common bovine contaminants from cell culture and affinity columns, various proteolytic enzymes, affinity tags, and other contaminants. When compared to the MaxQuant and cRAP contaminant lists, our new FASTA library is up-todate for all Uniprot IDs and contains an additional 166 contaminant proteins. This new FASTA library can be used for both DDA and DIA proteomics. We also added a "Cont" prefix in each contaminant entry in the FASTA library, allowing contaminant proteins to be easily filtered and removed in the result files.

## 2. Removing Contaminant Proteins from Result Files.

- 2.1. Launch the results file in Microsoft Excel. In the "Home" tab, click on "Sort & Filter" and then "Filter".
- 2.2. Navigate to the Protein ID column and type in "Cont".
- 2.3. This will select all contaminant proteins. All contaminant proteins should be removed prior to statistical analysis.

# 3. Including a Contaminant FASTA library in Proteome Discoverer DDA Workflows

3.1. Click the "Administration" tab and select "Maintain Fasta Files". Click "Add" and then select "Protein Contaminants\_Hao Lab.Fasta".

e View Administration Tools Window Help			» ▼	<i>4</i>	6	<b>,</b> >
Sart Page × Administration ×			» ▼	Ŕ		<b>R</b> >
Start Page X Administration X						- min
						▼ 4 ⊅
	📫 Add 🎬 Download 🎬 Undate 🤰	Remove 🔊 Can	rel 🛄 Ex	ort 2	Refresh	×
Process Management *	Name	Protei Tavo Versio	n Eila Siz # G	an # Pa	Qt Lact N	• Undate
- A	PCCA human facta	Curt	0	1 779	Au 02/1	
Job Queue	EndoBiotinulated carboxulase Humanfasta	Cust	0	7 8537	Av 02/1	
	24Mitochondrial derived pentides fasta	Cust	3	24 1472	Av 02/2	
Content Management *	257 Human Neuropeotide unreviewed re	Cust	103 3	39 801	Av 02/2	
and the second s	377Lysosome_human.fasta	Cust	225 3	36 193	Av 03/2	
FASTA Files	PCCA mouse,fasta	Cust	0	1 724	Av 03/2	
EASTA Indexes	533 Lysosome_Mousefasta	Cust	314 5	19 265	Av 03/2	
Z F PASTA LIGEXES	Streptavidin.fasta	Cust	0	1 183	Av 04/2	
EASTA Parsing Rules	Mouse_Swissprotfasta	Cust	11303 17	0 967	Av 06/2	
S	Human_Swissprotfasta	Cust	13290 20	3 113	Av 06/2	
Spectral Libraries	HRP.fasta	Cust	0	1 353	Av 07/1	
	BSA.fasta	Cust	0	1 607	Av 10/2	
Chemical Modifications	Human LAMP1.fasta	Cust	0	1 417	Av 07/1	
	218Mouse_neuropeptidefasta	Cust	91 1	.97 717	Av 08/0	
Cleavage Reagents	Mouse_NPFF.fasta	Cust	5	15 4247	Av 08/0	
	Human_SwissProt_20375.fasta	Cust	13291 20	3 113	Av 01/0	
Annotation Aspeds	Human_Mitochondria_reviewed_1200.fasta	Cust	802 1	75 676	Av 01/0	
	Mouse_Mitochondrion_1839fasta	Cust	1032 18	36 848	Av 01/1	
Quantification Methods	DELE1.fasta	Cust	2	4 2041	Av 01/1	
	Contamination.fasta	Cust	183 3	79 147	Av 02/0	
License Management	🕨 Protein Contaminants_Hao Lab.fasta 🛛 🤌	Cust	183 3	79 147	Av 02/0	

3.2. Open a new study and select a processing step workflow. Click on the "Sequest HT" tab. For protein database, select both the "Protein Contaminants\_Hao Lab" and organism FASTA for your sample.

**NOTE:** The protein contaminant FASTA file must be included to ensure the algorithm does not misassign peptides to the wrong protein.



3.3. Select your consensus step workflow. Under the "Protein Marker" tab, select a contaminant database. This will create a separate column in the result file marking contaminant proteins.

Parameters of 'Protein Marker'	👫 Open 📸 Open Common 🛔 Save 🔣 Save Common 👫 Auto Layout 🐹 Clear
Show Advanced Parameters	Wolkflow: CWF, Comprehensive, Enhanced Annotation, LFQ, and Precursor, Quan, HEK DefaultCharge
	Volte: Unit_Control the testing of the labor in the action of the second rest outer the definition of the second rest of the s
Protein Database A database containing potential contaminations to mark in the result.	Protein Annotation 8 Protein FDB 7 Protein Grouping 5

## 4. Including a Contaminant FASTA library in a DDA MaxQuant Workflow

- 4.1. Launch MaxQuant. Load .*raw* files. Click the "Global parameters" tab and then select "Sequences".
- 4.2. Select the "Protein Contaminants\_Hao Lab.Fasta" and then click on "Identifier rule".
- 4.3. Unselect "Include contaminants".

**NOTE:** Including the MaxQuant contaminant database will not affect results. However, this database includes UniProt IDs that have since been removed or reassigned.

Sequences Protein quantification Tables	MS/MS analyze	r Advan	ce visualization	Configuration						
Identification Label free quantification Fold Parameter	erlocations MS section	S/MS frag	mentation							
Fasta files	Add Remove Change folder Variation rule Test		Change folder	Identifier rule Description		Taxonomy rule Taxonomy ID				
		Fasta file path						Identifier rule	Description rule	
	1	C.IU	C.\Users\HaoLab\Desktop\Human_Swissprot.fasta				True	>([^\S]*)	>(.*)	
	2	C:W	sers\HaoLab\De	sktop/Protein Co	ntaminants_I	lao Lab fasta	True >([^\s]*)	>([^\(g]*)	>(*)	
nclude contaminants Min. peptide length	7				>."\(.")\	Parse rule	Descrip	ption		
Max. peptide mass [Da]	4600				1	>.*V(.*)V	UniProt identifier			
Min. peptide length for unspecific search	8				2	>(giV[0-9]*)	NCBI accession			
Max, peptide length for unspecific search	25				3	>IPI:([^\].]*)	IPI acci	ession		
Variation mode	None				4	>(.*)	Everyth	ning after ">"		
					5	>([^]*)	Up to fi	irst space		
Proteogenomics fasta files	Ad	d	Remove	Change folder	6	>([~[t]*)	Up to fi	irst tab character		
	Te	st			1					
		Fast	ta file path						T e	
					6 items	1			100 % ~ 1	
					Cancel				100 %	

## 5. Including a Contaminant FASTA library in a MaxDIA Workflow

5.1. For library-based DIA proteomics, you must include the same contaminant and species specific FASTA files used to generate the spectral library. These FASTA files will be included following Steps 2-3.

## 6. Integrating a Contaminant FASTA library into a Library-based Biognosys Spectronaut Workflow

- 6.1. Launch Biogenesis Spectronaut and select the "Databases" tab. Import the "Protein Contaminants\_Hao Lab.Fasta".
- 6.2. Select the "Library" tab. Click "Generate Library from Pulsar/Search Archives".
- 6.3. Select "Add Runs from File" to add .raw files.

**Note:** The *.raw* files from our custom contaminant-only experiment can be included to ensure the accurate detection and inclusion of contaminant spectra within the library.

Library	Analysis	Post Analysis	Report	QC	Pipeline	Databases	Settings	About	
Spectral	Libraries gnosys Spect Files, 2 Folders - ectral Libraries des, 8 Folders August 2021 0 Files, 0 Folder December 20 0 Files, 0 Folder December 20	Search Archives ral Library Online. Right Click to Down s rs pectral Libraries rs 221	E	Set up Choose a HEK/HeLa A <> HEK	D Library Gener an experiment r a Contaminant L (VHeLa Contam [01] 20210813 [02] 20210813 [03] 20210813 [04] 202108	ration from Pul lame and select library _Contamination _Contamination _Contamination _Contamination _Contamination 13_Contamination	sar the raw files y _Study_1ug_Sr _Study_1ug_Tr _Study_1ug_Tr on_Study_GE_	eqGradeTry rypsin_Gold rypsinLysC. MagBeads_1	earch.
Search 2. Generat Export S Import S	e Library from te Library from spectral Library	Pulsar / Search Arc n Proteome Discor	+ hives. rerer 3	Î Î Î Î Î Î Î Î Î	(05) 202108 (06) 202108 (07) Fig4_HEk (08) Fig4_HEk (09) Fig4_HEk (10) Fig4_HEk (11) Fig4_HEk (12) Fig4_HEk (12) Fig4_HEk	13_Contamination 13_Contamination 13_Contamination 13_Contamination 13_20-1 m_DDA_f 1293-1 m_DDA_f 1293-1 m_DDA_f 1293-1 m_HPRP- 111 111 Remove	on_Study_Sign on_Study_Sign R01_T0.raw R02_T0.raw R03_T0.raw 5perc_DDA_R1 10perc_DDA_F 15perc_DDA_F	na_FlagBead na_HABeads 01_T0.raw R01_T0.raw R01_T0.raw	L_lug_Tr

6.4. Click "Next" and then "Fasta File." Select the "Protein Contaminantion\_Hao Lab FASTA". Select the remaining settings to build the desired library.



6.5. For library-based DIA proteomics, select the library that was built during data analysis and include the appropriate databases.

#### 7. Integrating a Contaminant FASTA library into a DIA-NN Workflow

- 7.1. Launch DIA-NN. Click "spectral library" and add the contaminant library that was built using Spectronaut.
- 7.2. Under "Add FASTA" select the appropriate FASTA libraries that were used to build the spectral library.

Input		
Raw diaPASE	F.d Clearlist	Convert to .dia
		^
Spectral library	C:\Ashlev\DIANN	LF HEK Cont Pn
	C:\Ashley	
AUGEASTA	\DIANN_LF_HEK_	Cont_Prefix
Clear list	Ashley	
Reannotate	\DIANN_LF_HEK_ \Human_Swissprot	Cont_Prefix .fasta
		×
DIA-NN exe	diann.exe	

#### 8. Including a Contaminant FASTA library into a Skyline Workflow

- 8.1. Launch Skyline (version 21.2) and open a "Blank Document".
- 8.2. A spectral library can be built by selecting "File", "Import" and then "Peptide Search."
- 8.3. Import the *.pdResult* file from Proteome Discoverer or *msms.text* file from MaxQuant. Select "Next" to build the process of building the peptide search library.

1 Import Peptide Search	×	
Spectral Library		
Build Use existing		
Cut-off score: 0.95		
Start from: Search results (build library directly)	~	
Result files: DDA_Mouse_Cortex_Injection1.pdRet	Add Files Remove Files	
iRT standard peptides:	1 Building Peptide Search Library	×
None     ✓       Include ambiguous matches     Filter for document peptides	Parsing 531574 spectra.	
Workflow  DDA with MS1 filtering  DIA		Cancel
O PRM	n Next > Cancel	

- 8.4. Select the appropriate .raw files and click "Next".
- 8.5. Select the FASTA File and then "Finish".

**NOTE:** Only a single FASTA library can be imported. The contaminant FASTA file will need to be combined with the organism FASTA.

import optice search		
Import FASTA (o	ptional)	
Enzyme:	Max missed cleavages:	
Trypsin [KR   P]	~ 0 ~	
FASTA records begin with '> optional protein description.	and have the protein name followed	by the
		Browse

8.6. Library-based DIA analysis can be conducted using established Skyline workflows. However, the conjoined FASTA file used to build the library should be included during data analysis.

### 9. Including a Contaminant FASTA library into a PECAN Workflow

- 9.1. Launch EncylopeDIA (version 1.12.31). Select the Walnut tab.
- 9.2. Import the contaminant FASTA library to the "Background" and "Target" sections.

**NOTE:** Only a single FASTA library can be imported into the workflow. The Hao Lab Contaminant library must be combined with your organism FASTA database.

File View Convert Data Hel	p	
EncyclopeDIA	s 🚳 Walnut	
Walne Direc (DIA) Walnut u: chromato features.	ut: PeCAn-based Peptide Detecti tly from Data-Independent Acquis MS/MS Data ses PeCAn-style scoring to extract peptide fragmentation grams from MZML files, assign peaks, and calculate vari These features are interpreted by Percolator to identify p	on sition <sup>n</sup> ous peak eptides.
Parameters:	Partie Castroineth Under De Richard	<b>5</b> 44
Background supplemental FASTA	Protein Contaminants_Hao Lab_Pretix.rasta	Edit
Target: Supplemental FASTA Protein	Contaminants_Hao Lab_Prefix.fasta	Edit
Target/Decoy Approach: Normal	Target/Decoy	~
Precursor Window Width (blan	k=extract from file): -1	
Enzyme: Trypsin		~
Fixed: C+57 (Carbamidomethyl)		~
Fragmentation: CID/HCD (B/Y)		~
Precursor Mass Tolerance: 10.	U PPM	~
Fragment Mass Tolerance: 10.0	D PPM	~
Maximum Missed Cleavage:		2 🌻
Percolator Version: v3-01		~
Number of Quantitative lons:		5
Number of Cores:		6 🌻
Charge range:	2 🜩 to 4 🖨	
Additonal Command Line Optic	ons:	

EncyclopeDIA Graphical Interface (version 1.12.31)