# Step-by-Step Tutorial for Integrating Contaminant FASTA and Spectral Libraries in Various DDA and DIA Proteomics Software Platforms

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*. Here, we generated universal contaminant FASTA and spectral libraries that can be used for both data-dependent acquisition (DDA) and data-independent acquisition (DIA) proteomics, available to download at: <u>https://github.com/HaoGroup-ProtContLib</u>, and <u>ProteomeXchange</u> (#PXD031139) 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 modified the contaminant FASTA library to contain a "Cont" prefix before each UniProt identifier, simplifying the process of searching and removing contaminant proteins prior to statistical analysis.

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

#### Please cite our publication:

Ashley M. Frankenfield, Jiawei Ni, Mustafa Ahmed, Ling Hao, "Protein Contaminants Matter: Building Universal Protein Contaminant Libraries for DDA and DIA Proteomics", Journal of Proteome Research, 2022. <u>https://doi.org/10.1021/acs.jproteome.2c00145</u>

(PDF is available on <u>BioRxiv</u> and the Hao Lab Website: <u>https://blogs.gwu.edu/haolab/research/publication/</u>)

# Table of Content:

- 1. <u>Brief Description of Contaminant Libraries</u>
- 2. <u>Removing Protein Contaminants from Result File in Excel</u>
- 3. Using Contaminant FASTA in DDA Software
  - 3.1 Proteome Discoverer for DDA
  - 3.2 MaxQuant for DDA
- 4. <u>Building Contaminant Protein Spectral Libraries</u>
  - 4.1 Building a Spectral Library in Spectronaut
  - 4.2 Building a Spectral Library in MaxQuant
- 5. Using Contaminant FASTA and Spectral Libraries in Library-based DIA
  - 5.1 Spectronaut for library-based DIA
  - 5.2 DIA-NN for library-based DIA
  - 5.3 Skyline for library-based DIA
  - 5.4 MaxDIA for library-based DIA
- 6. Using Contaminant FASTA in Library-free DIA
  - 6.1 Spectronaut for library-free DIA
  - 6.2 DIA-NN for library-free DIA
  - 6.3 PECAN for library-free DIA

## 1. Brief Description of Contaminant Libraries

Exogenous contaminant proteins originated from reagents and sample handling are often shared in most 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, containing deleted Uniprot IDs, sample-specific interference proteins that are incorrectly listed as contaminants, and commercially available human protein standards 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 for 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-to-date 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.

- 1.1. Launch the results file in Microsoft Excel. In the "Home" tab, click on "Sort & Filter" and then "Filter".
- 1.2. Navigate to the Protein ID column and type in "Cont\_".

	a 5-	¢					Report_Sc	heduled DIA	(Normal) -	Excel			Sign in		- (	- x
F	ile Ho	ome In	sert Page	Layout	Formulas	Data	Review	View H	elp Acro	obat 💡	Tell me what	t you want t	o do			A Share
Pa	ste 🛷	Calibri B I I	-   11 1 -   🖽 - Font	▲ - <u>▲</u>	· = =	Alignment	€. €.	Seneral \$ - % * Numbe	+0 00 +0 00	Conditiona Formatting	al Format as • Table • Styles	Cell Styles +	Format • Cells	∑ A Z Sc Fi E	The select	2
E1				• 1 0	~ ~	fx PG.F	ProteinDe	scriptions								^
		P		0	c		c		1	1 1	r		м	N	0	
1	P Cond T	DG Eard	PG Gort T	PG Prof T	G Prot a	DED Str -	DED MS -		EG Dati z	FG Ova z	EG Ano X	EG Sual y	EG Dual T	EG Sign x	EG Harl Y	EG Sign
203	Not Defin	Protein	Contaminat	Cont P00	Trunsin	I GEHNID	3 225+08	71611928	5 639311	1 725-08	84 5584	0.00026	A 275-07	2781 781	FAISE	2781 7
204	Not Defin	Protein	Contaminat	Cont POO	Trypsin	LGEHNIDY	3.22E+08	71611928	4 226716	3.77E-08	84 51532	0.656757	2 38F-05	5 37469	TRUE	5.374
205	Not Defin	Protein	Contaminat	Cont POO	Trypsin	VATVSI PR	31821202	9959710	4.676975	1 725-08	49 5358	0.001594	7 34F-07	182 2330	FAISE	182.22
206	Not Defin	Protein	Contaminat	Cont P00	Trypsin	VATVSLPR	31821202	9959710	5.026395	2.76E-07	49.47039	0.956123	0.000253	138.0881	TRUE	138.05
207	Not Defin	Protein	Contaminat	i Cont P00	Trypsin	IITHPNEN	1.3E+08	9900832	4.626133	6.17E-08	88.75113	0.723467	4.29E-05	135.0102	FALSE	135.01
208	Not Defin	Protein	Contaminat	Cont P00	Trypsin	LSSPATLN	2956277	1024684	5.171885	1.72E-08	38,64737	0.118914	2.99E-06	100.8617	FALSE	100.86
209	Not Defin	Protein	Contaminat	Cont POO	Trypsin	LGEHNIDV	5093644	464217.3	4,667501	1.72E-08	116.8189	0.032981	1.91E-06	105,7796	FALSE	105.77
210	Not Defin	Protein	Contaminat	Cont POO	Trynsin	ITHPNEN	1625918	79036 19	5 399454	1.91F-08	99.433	0.4792	8 78F-06	22.01384	TRUF	22 013

1.3. This will select all contaminant proteins. Evaluate the selected proteins to ensure that they are not biologically relevant based on custom sample types. Remove contaminant proteins prior to statistical analysis.

## 3. Using Contaminant FASTA in DDA Software

#### 3.1. Proteome Discoverer for DDA

3.1.1. Click the "Administration" tab and select "Maintain Fasta Files". Click "Add" and then select "Protein Contaminants Hao Lab.Fasta".

M Thermo Proteome Discoverer 2.4.1.15									_		;	<
File View Administration Tools Window	Help											
💱 🚮 🍯 🔒 🙄 🗖	1 🌱				>: •	}		Ŕ	and the second	5 (		>
Start Page × Administration ×											- 4	
Add 👹 Download 🚿 Update 🧩 Remove 🚫 Cancel 🔛 Export 🥭 Refresh 🐥												
Process Management	* <b>-</b>	Name	Protei	Taxo	Version	File Siz	# Sea	# Re	9	Last M	Undate	Ē
A lab Queue		PCCA human fasta	Cust	Tuxo	Version	0	1	728	Δv	02/1		
Since And Age		EndoBiotinylated carboxylase Humanfasta	Cust			9	7	8537	Av	02/1		
	_	24Mitochondrial derived pentides fasta	Cust			3	24	1472	Av	02/2		
Content Managenient	*	257 Human Neuropeptide unreviewed re	Cust			103	239	801	Av	02/2	Π	
Contraction and the second sec		377 Lysosome human.fasta	Cust			225	336	193	Av	03/2	Γ	
FASTA Files		PCCA_mouse.fasta	Cust			0	1	724	Av	03/2	Г	
EASTA Indexes		533 Lysosome_Mousefasta	Cust			314	519	265	Αv	03/2		
		Streptavidin.fasta	Cust			0	1	183	Av	04/2		
FASTA Parsing Rules		Mouse_Swissprotfasta	Cust			11303	170	967	A٧	06/2		
S		Human_Swissprot <i>fa</i> sta	Cust			13290	203	113	Αv	06/2		
Spectral Libraries		HRP.fasta	Cust			0	1	353	A٧	07/1		
CITTER .		BSA.fasta	Cust			0	1	607	Αv	10/2		
Chemical Modifications		Human LAMP1.fasta	Cust			0	1	417	Αv	07/1		
		218Mouse_neuropeptidefasta	Cust			91	197	717	Αv	08/0		
Cleavage Reagents		Mouse_NPFF.fasta	Cust			5	15	4247	Αv	08/0		
		Human_SwissProt_20375.fasta	Cust			13291	203	113	A٧	01/0		
Annotation Aspeds		Human_Mitochondria_reviewed_1200.fasta	Cust			802	1175	676	A٧	01/0		
		Mouse_Mitochondrion_1839.fasta	Cust			1032	1836	848	A٧	01/1		
Quantification Methods		DELE1.fasta	Cust			2	4	2041	A٧	01/1		
		Contamination.fasta	Cust			183	379	147	Αv	02/0		
License Management	*	Protein Contaminants_Hao Lab.fasta 🛛 🖉	Cust			183	379	147	Av	02/0		į.

3.1.2. Click on the "Sequest HT" tab in the processing workflow in a study file. 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 in data processing step to reduce protein/peptide false identifications.



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



3.1.4. Contaminant proteins can be filtered using the accession column or separate contaminant column.

	Prote	ins 😴	Protein Grou	ps P	eptide Grou	Ips PSMs	MS/MS Spectrum Info	Input Files	8	Specializ	ed Traces	Consensus Featur
	₽.	Checked	Protein FDI +	Master	Accession	Description			Ex	p. q-valu 🕂	Contaminant	Sum PEP Score -
1	-12		High	V	Cont_P00	Fructose-bisph	osphate aldolase A OS=Ory	ctolagus cunicu	h	0.000	X	252.038
2	-12		High	*	Cont_P02	Albumin OS=B	los taurus OX=9913 GN=AL	LB PE=1 SV=4		0.000	Х	237.536
	-12		High	$\checkmark$	Cont_Q32	Glucose-6-pho	Glucose-6-phosphate isomerase OS=Bos taurus OX=9913				Х	136.760
4	-		High	~	Cont_P00	Trypsin OS=Sus scrofa OX=9823 PE=1 SV=1				0.000	Х	127.046
Ę	; -¤		High	×	Cont_P05	Keratin, type I	Keratin, type I cytoskeletal 18 OS=Mus musculus OX=1009(			0.000	Х	98.765
6	-		High	~	Cont_Q5ł	Isoform 2 of Tr	opomyosin beta chain OS=B	os taurus OX=	YE .	0.000	Х	89.594
7	-		High	1	Cont_E1E	Tubulin beta d	ubulin beta chain OS=Bos taurus OX=9913 GN=TUBB1 P 0.000		х	55.926		
8	-		High	V	Cont_P02	202 Profilin-1 OS=Bos taurus OX=9913 GN=PFN1 PE=1 SV=2 0.00		0.000	Х	35.402		
9			High	1	Cont_P12	Alpha-2-HS-gl	vcoprotein OS=Bos taurus O	X=9913 GN=A	ł:	0.000	X	11.762
	0 🕁		High	×	Cont_P34	Alpha-1-antipro	oteinase OS=Bos taurus OX:	=9913 GN=SEF	F	0.000	X	11.306
1	1 👳		High	V	Cont_Q35	Alpha-1-acid g	lycoprotein OS=Bos taurus (	X=9913 GN=C	F	0.000	Х	5.492
1	2 🗇		High	4	Cont_Q6I	Keratin, type II	Keratin, type II cytoskeletal 72 OS=Mus musculus OX=1005 0.001				Х	4.727
1	3 -=		High	V	Cont P22	Streptavidin O	S=Streptomyces avidinii OX=	=1895 PE=1 SV	-	0.007	Х	2.800

### 3.2. MaxQuant for DDA

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

**NOTE:** Including the existing MaxQuant contaminant database will not affect results. However, contaminant proteins from the new FASTA will not be marked in the contaminant column in the MaxQuant results file, which may lead to confusion. Contaminant proteins are marked in the UniProt ID column with the prefix "Cont\_" as described on page 2.

Haw data Group-specific parameters Globy	al parameters Per	rformance V	Isualization Co	onfiguration						
Sequences Protein quantification Tables	MS/MS analyzer	Advanced								
Identification Label free quantification Fold Parameter :	erlocations MS/I section	MS fragmentat	tion							
Fasta files	Add	1	Remove Change folder			e Description	Taxor	nomy rule	omy rule Taxonomy	
	Variation	rule	Test				and first-street	the second s		
		Fasta file	path				Exists	Identifier r	ule	Description rule
	1	C:\Users\	HaoLab\Desk	top\Human_S	wissprot.fast	а	True	>([^\s]*)		>(.*)
	2	C:\Users\	HaoLab\Desk	top/Protein Co	ntaminants_	Hao Lab fasta	True	>([^\s]*)		>(.*)
Include contaminants					-					
Include contaminants		selected			>."\(.")\					
Include contaminants Min. peptide length	7	selected			>."\(.")\	Parse rule	Descrip	otion		
Include contaminants Min. peptide length Max. peptide mass [Da]	7	selected			>.*\(.*)\  1	Parse rule	Descrip	ption Lidentifier		
Include contaminants Min. peptide length Max. peptide mass [Da] Min. peptide length for unspecific search	7 4600 8	abecieu			>."\(.")\  1 2	Parse rule *\(.*)\ >(gi\[[0-9]*)	Descrip UniPro	otion tidentifier ccession		
Include contaminants Mn. peptide length Max, peptide mass [Da] Mn. peptide length for unspecific search Max, peptide length for unspecific search	7 4600 8 25				>."\(,")\  1 2 3	Parse rule >*\\(.*)\\ >(gi\[0-9]*) >IPI:([^\].]*)	Descrip UniPro NCBI a	otion t identifier ccession ession		
Include contaminants Mn. peptide length Max, peptide mass [Da] Mn. peptide length for unspecific search Max, peptide length for unspecific search Variation mode	7 4600 8 25 None				>."\(.")\  1 2 3 4	Parse rule >.\\(.*)\ >(gi\[[0-9]*) >IPI:([^\], ]*) >(.*)	Descrip UniPro NCBI a IPI acco Everyth	otion tidentifier ccession ession hing after ">"		
Include contaminants Mn. peptide length Max, peptide mass [Da] Mn. peptide length for unspecific search Max, peptide length for unspecific search Variation mode	7 4600 8 25 None				>."\(.")\  1 2 3 4 5	Parse rule           >.*\(.*)\           >(gi\(0-9]*)           >IPI:([^\\].]*)           >(.*)           >([^\].]*)	Descrip UniPro NCBI a IPI acco Everyth Up to fi	tidentifier ccession ession ning after ">"	R	
Include contaminants Mn. peptide length Max, peptide mass [Da] Mn. peptide length for unspecific search Max, peptide length for unspecific search Variation mode Proteogenomics fasta files	7 4600 8 25 None		Remove	Change folder	>."\(.')\  1 2 3 4 5 6	Parse rule           >*\\(.*)\\           >(gi\[0-9]*)           >IFI:([^N\], ]*)           >(.*)           >([^{^1})*)           >([^{^1}]*)	Descrip UniPro NCBI a IPI acco Everyth Up to fi	t identifier ccession ession hing after ">" irst space irst tab chara	acter	
Include contaminants Min. peptide length Max, peptide mass [Da] Min. peptide length for unspecific search Max, peptide length for unspecific search Variation mode Proteogenomics fasta files	7 4600 8 25 None Add		Remove	Change folder	>.*\(.')\  1 2 3 4 5 6	Parse rule           >.*\(.*)\(           >(g)\([0.9]*)           > P1.(?u ]*)           >(.*)           >([^1]*)           >([^1]*)	Descrip UniPro NCBI a IPI acco Everyth Up to fi	btion t identifier ccession ession hing after ">" irst space irst tab chara	acter	
Include contaminants Min. peptide length Max, peptide mass [Da] Min. peptide length for unspecific search Max, peptide length for unspecific search Variation mode Proteogenomics fasta files	7 4600 8 25 None Add Test	Fasta file	Remove	Change folder	>.*\(.')\  1 2 3 4 5 6	Parse rule           >*\(\(^*)\)           >(g)\([0-9]*)           >IP1:([^NL], ]*)           >(.*)           >([^N]*)           >([^N]*)	Descrip UniPro NCBI a IPI acco Everyth Up to fi Up to fi	t identifier ccession ession ning after ">" irst space irst tab chara	acter	
Include contaminants Min. peptide length Max, peptide mass [Da] Min. peptide length for unspecific search Max, peptide length for unspecific search Variation mode Proteogenomics fasta files	7 4600 8 25 None Add Test	Fasta file	Remove	Change folder	>:*\('.')\ 1 2 3 4 5 6	Parse rule           >*\(\(.')\)           >(g)\([0-9]*)           >IPL([0^1, ]*)           >(.*)           >([^1]*)           >([^1]*)	Descrip UniPro NCBI a IPI aco Everyth Up to fi	bion tidentifier ccession ession ning after ">" rst space rst tab chara	acter	

## 4. Building Contaminant Protein Spectral Libraries

To establish comprehensive contaminant protein spectral libraries for DIA proteomics, we created a series of contaminant-only samples using various proteolytic enzymes, affinity purification beads and fetal bovine serum (FBS) that are commonly used for cell culture medium. Contaminant Protein Spectral Library is available to download in Github and ProteomeXchange (#PXD031139). For proteomics software that allows the input of multiple spectral libraries, our contaminant spectral library and custom proteomics data can be included together. For software that only allows one spectral library input, an integrated spectral library can be built using our contaminant-only raw data and custom proteomics data. We have tested that the integrated spectral library performs similarly to two separate libraries. Either method is better compared to the results analyzed without the contaminant library.

### 4.1. Building a Spectral Library in Spectronaut

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

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

Library Analysis Post Analysis Repo	rt QC Pipeline Databases Settings About
Spectral Libraries         Search Archives           Image: Spectral Library Online         Image: Spectral Library Online           Image: Spectral Libraries         Spectral Libraries	Set up Library Generation from Pulsar      Choose an experiment name and select the raw files you want to search.
12 Files, 8 Folders August 2021 ○ Files, 0 Folders December 2021 ○ Files, 0 Folders December 2021 ○ Files, 0 Folders February 2022 ○ Files, 0 Folders January 2022 ○ Files, 0 Folders December 2021 ○ Files, 0 Folders December 2021 O Files, 0 Folders December 2021 O Files, 0 F	HEK/HeLa Contaminant Library [01] 20210813_Contamination_Study_1ug_SeqGradeTrypsin.raw [02] 20210813_Contamination_Study_1ug_Trypsin_Gold.raw [03] 20210813_Contamination_Study_1ug_Trypsin_LysC.raw [04] 20210813_Contamination_Study_GE_MagBeads_1ug_Tryp [05] 20210813_Contamination_Study_Sigma_FlagBead_1ug_Tr [06] 20210813_Contamination_Study_Sigma_HABeads_1ug_Tr
O Flies, 0 Folders     O Flies, 0 Folders     Search      Generate Library from Pulsar / Search Archives.     Generate Library from Proteome Discoverer     Export Spectral Library     Import Spectral Library	[07] Fig4_HEK293-1m_DDA_R01_T0.raw     [08] Fig4_HEK293-1m_DDA_R02_T0.raw     [09] Fig4_HEK293-1m_DDA_R03_T0.raw     [10] Fig4_HEK293-1m-HPRP-5perc_DDA_R01_T0.raw     [11] Fig4_HEK293-1m-HPRP-10perc_DDA_R01_T0.raw     [12] Fig4_HEK293-1m-HPRP-15perc_DDA_R01_T0.raw     [12] Fig4_HEK293-1m-HPRP-15perc_DDA_R01_T0.raw     [12] Fig4_HEK293-1m_HPRP-15perc_DDA_R01_T0.raw     [12] Fig4_HEK293-1m_HPRP-15perc_DDA_R01_T0.raw     [12] Fig4_HEK293-1m_HPRP-15perc_DDA_R01_T0.raw     [12] Fig4_HEK293-1m_HPRP-15perc_DDA_R01_T0.raw     [12] Fig4_HEK293-1m_HPRP-15perc_DDA_R01_T0.raw     [12] Fig4_HEK293-1m_HPRP-15perc_DDA_R01_T0.raw     [12] Fig4_HEK293-1m_HPRP-15perc_DDA_R01_T0.raw

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

Library Analysis Post Analysis Report		Pipeline Data	bases Settings	About
Choose Fasta File(s)			- 0	×
Spectral Librar       From Recent         Biognosy       From Recent         Spectral       From Recent         Spectral       From Recent         12 Flas, 8       From Recent         Augu       From Recent         Decc       Statistics         Decc       Contamination_Update         Decc       SwissProt_Ren         25,390 Entries       SwissProt All Organism         Softer       SwissProt All Organism         Softer       SwissProt All Organism         Offer       SwissProt All Organism         Offer       SwissProt All Organism         Offer       Contaminants	13 s_Hao Lab ed_FastaFil_ viewed_Ca_ ms iewed	Entries: Date Created: Date Modified: Organism: Protein Id: Description:	lame:	ment.
Nove 0 File: 0 File	13	-		TO raw TO raw TO raw TO raw QK TO raw

## 4.2. Building a Spectral Library in MaxQuant

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

**NOTE:** Including the existing MaxQuant contaminant database will not affect results. However, contaminant proteins from the new FASTA will not be marked in the contaminant column in the MaxQuant results file, which may lead to confusion. Contaminant proteins are marked in the UniProt ID column with the prefix "Cont\_" as described on page 2.

Raw data Group-specific parameters Glob	MS/MS apply and	ormance Visuali:	zation Co	onfiguration							
Identification Label free quantification Fold Parameter	er locations MS/N section	IS fragmentation									
Fasta files	Add	Remo	ove	Change folder	Identifier ru	e Description	Taxo	nomy rule Taxon	omy ID		
	Variation r	ule Tes	st		-						
		Fasta file path	1				Exists	Identifier rule	Desc	ription	n rule
	1	C:\Users\Haol	Lab\Desk	top\Human_S	wissprot.fast	а	True	>([^\s]*)	>(.*)		
	2	C:\Users\Haol	Lab\Desk	top/Protein Co	ontaminants_	Hao Lab.fasta	True	>([^\s]*)	>(*)		
Min. peptide length Max. peptide mass [Da]	7				1	Parse rule	Descrip	ption Lidentifier			
wax, pepude mass [Ua]	4600				2	1         >         1           2         >(gi\[0-9]*)         3         >IP1:([^N].J*)           3         >IP1:([^N].J*)         4         >(.*)           5         >([^{^1})*)         >(.*)		NCBI accession IPI accession Everything after ">*			_
Max, peptide length for unspecific search	8				3						_
Variation mode	25				4						-
	None				5			irst space			
Protocomo fonto filmo					6	>([^\t]*)	Up to f	irst tab character			
Proteogenomics fasta files	Add	Remo	ove	Change folder							
	Test	Fasta file path	i.								Т
	-					31		-	10		
					6 items				100 % 🗸	1	
					Cancel				0	K	

### 5. Using Contaminant FASTA and Spectral Libraries in Library-based DIA

#### 5.1. Spectronaut for Library-based DIA.

5.1.1. Launch Biognosys Spectronaut. Select "Set up DIA Analysis from File".

Spectro	naut								_	$\times$
	$\mathbf{A}$	00	<b>~~</b>	<b></b>	<b>a</b>		<b>(</b> )	ĺ		
Library	Analysis	Post Analysis	Report	QC	Pipeline	Databases	Settings	About		
Load Exp	eriment ×	¢								$\mathbf{x} \neq \mathbf{x}$
<ul> <li>Set up Start a</li> <li>Set up Start a</li> </ul>	a DIA Analysis classic DIA ar a directDIA <sup>TT</sup> A library-free DI Spectronaut E	s from File halysis using a tar <u>malysis from File</u> A analysis using a speriment	geted, library  a protein data	r-based search. base (FASTA) fi	ile as input.					
Open a	a previously sa nt Experiments	ved Spectronaut	experiment fr	om a *.SNE file.						

5.1.2. Load the *.raw* files for the study.

Set up DIA Analysis			$\times$
Specify all LC-MS/MS measurements that you want to include in this experiment and assign which spectral-libraries to Each run needs to have at least one spectral-library assigned in order to proceed.	use.		
Injection 1			
<ul> <li>Injection1         <ul> <li>[1] 20210824_DIA_HEK_Injection1.raw</li> <li>[2] 20210824_DIA_HEK_Injection2.raw</li> <li>[3] 20210824_DIA_HEK_Injection3.raw</li> <li>[4] 20210824_DIA_HEK_Injection4.raw</li> </ul> </li> </ul>			
Add Runs from File     Assign Spectral Library     Remove	Next	Skip to L	_ast

5.1.3. Select the contaminant-containing spectral library and contaminant FASTA used during library creation.

#### 5.2. DIA-NN for Library-based DIA.

- 5.2.1. Launch DIA-NN. Click "Spectral library" and add the contaminant FASTA that was built using Spectronaut.
- 5.2.2. Load the *.raw* files. Under "Add FASTA" select the appropriate FASTA libraries to build the spectral library.

Input	
Raw diaPASE	F.d Clear list Convert to .dia
	^
	~
Spectral library	C:\Ashley\DIANN_LF_HEK_Cont_Pn
Add FASTA	C:\Ashley
Clear list	\ContaminationPrefix.fastaC: \Ashley
Reannotate	\DIANN_LF_HEK_Cont_Prefix \Human_Swissprot.fasta
DIA-NN exe	diann.exe

### 5.3. Skyline for Library-based DIA.

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

<u>1</u> 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.pdRe	Add Files Remove Files	
iRT standard peptides:	🗽 Building Peptide Search Library	×
None   Include ambiguous matches  Fitter for document peptides  Workflow  DDA with MS1 filtering  DDA	Pansing 531574 spectra.	Cancel
O PRM		
Finis	h Next > Cancel	

- 5.3.4. Select the appropriate .raw files and click "Next".
- 5.3.5. 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 Peptide Search	$\times$
Import FASTA (optional)	
Enzyme: Max missed cleavages: Trypsin [KR   P]	
FASTA records begin with '>' and have the protein name followed by the optional protein description.	ie vse

5.3.6. 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.

### 5.4. MaxDIA for Library-based DIA.

- 5.4.1. Launch MaxQuant. Load .raw files. Click the "Global parameters" tab and then select "Sequences".
- 5.4.2. For library-based DIA proteomics, you must include the same contaminant and organism specific FASTA files used to generate the spectral library. Select the "Protein Contaminants\_Hao Lab.Fasta" and the organism specific UniProt FASTA file. Click on "Identifier rule" and select "UniProt Identifier".
- 5.4.3. Select the "Group-Specific Parameters" tab. Click "Type" and select "MaxDIA" from the drop-down menu.
- 5.4.4. Import the peptide, evidence and msms.*txt* file for library-based DIA.

🔀 session1 - MaxQuant					$\times$
File Tools Window Help					
Raw data Group-specific parameters Global parameters Performance Visu	ualization Configuration				
Group 0 Type Modifications Instrument First search Digestion Label-free quartification Misc.					
Parameter group Parameter section					
Type MaxDIA					~
Library type	MaxQuant				~
	Peptide files	Add file(s)	Remove file		 
	Evidence files				 
	Evidence mes	Add file(s)	Remove file		 
	Msms files	Add file(s)	Remove file		 
		/ 66 110 (0)	ricinove nie		 

### 6. Using Contaminant FASTA in Library-free DIA

#### 6.1. Spectronaut for library-free DIA

6.1.1. Launch Spectronaut. Click "Set up a directDIA Analysis from File".

Spectro	naut								_	×
Library	Analysis	Post Analysis	Report		Pipeline	Databases	Settings	About		
Load Exp	eriment ×	0								 → x
<ul> <li>Set up Start a</li> <li>Set up</li> </ul>	a DIA Analysis classic DIA ar	s from File nalysis using a tai	rgeted, library	r-based search.		I				
Start a	library-free D	IA analysis using	a protein data	base (FASTA) f	ile as input.					
Load a Open a	Spectronaut E previously sa	operiment	experiment fr	om a *.SNE file.						
Recer	nt Experiments	1								

#### 6.1.2. Load the *.raw* files for the study.

Set up DIA Analysis			$\times$
Specify all LC-MS/MS measurements that you want to include in this experiment and assign which spectral-libraries to Each run needs to have at least one spectral-library assigned in order to proceed.	o use.		
Injection 1			
A Net Injection1			
<ul> <li>[1] 20210824_DIA_HEK_Injection1.raw</li> <li>[2] 20210824_DIA_HEK_Injection2.raw</li> <li>[3] 20210824_DIA_HEK_Injection3.raw</li> <li>[4] 20210824_DIA_HEK_Injection4.raw</li> </ul>			
✓ Add Runs from File Assign Spectral Library Remove			
	Next	Skip to Li	ast

6.1.3. Select the contaminant FASTA and organism FASTA.

#### 6.2. DIA-NN for library-free DIA

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

Input	
Raw diaPASE	F.d Clear list Convert to .dia
	^
	~
Spectral library	C:\Ashley\DIANN_LF_HEK_Cont_Pn
Add FASTA	C:\Ashley  \DIANN_LE_HEK_Cont_Prefix
Clear list	\ContaminationPrefix.fastaC: \Ashley
Reannotate	\DIANN_LF_HEK_Cont_Prefix \Human_Swissprot.fasta
	×
DIA-NN exe	diann.exe

#### 6.3. PECAN for library-free DIA

- 6.3.1. Launch EncylopeDIA (version 1.12.31). Select the Walnut tab.
- 6.3.2. 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 Help	
👿 EncyclopeDIA 🗽 Thesaurus 🧕 Walnut	
Walnut: PeCAn-based Peptide De Directly from Data-Independent A (DIA) MS/MS Data Walnut uses PeCAn-style scoring to extract peptide fragm chromatograms from MZML files, assign peaks, and calcu features. These features are interpreted by Percolator to in	etection cquisition nentation late various peak lentify peptides.
Parameters:	
Background Supplemental FASTA Protein Contaminants_Hao Lab_Prefix.fasta	Edit
Target: Supplemental FASTA Protein Contaminants_Hao Lab_Prefix.fasta	Edit
Target/Decoy Approach: Normal Target/Decoy	~
Precursor Window Width (blank=extract from file): -1	
Enzyme: Trypsin	~
Fixed: C+57 (Carbamidomethyl)	~
Fragmentation: CID/HCD (B/Y)	~
Precursor Mass Tolerance: 10.0 PPM	~
Fragment Mass Tolerance: 10.0 PPM	~
Maximum Missed Cleavage:	2 🔹
Percolator Version: v3-01	~
Number of Quantitative lons:	5 🔹
Number of Cores:	6 🔹
Charge range: 2 to 4 to	
Additonal Command Line Options:	
Consoler	

EncyclopeDIA Graphical Interface (version 1.12.31)