LINCS - Van Eyk lab
Reformat MS analysis output files for data release:

A. Un-normalized peptide data (Level 2)
B. Normalized peptide data (Level 3a)
C. Normalized protein data (Level 3b)

Center Name	Assay	Level0	Level1	Level2	Level3	Level4
	Conceptual		Raw data	Processed raw data	Normalized per sample data (can be used as input for statistical analysis)	Signatures (Can be used for "connecting" perturbations)
NeuroLINCS (proteomics)	SWATH Data independent acquisition	Raw data files (.wiff, .raw)	mzML, mzXML files	Non-normalized counts, limited QC based on FDR and scores (a) peptide counts	normalized counts, strict QC, including coeficient of variation (a) peptide (b) protein counts	Signatures, FOLD CHANGE (a) peptide level - skyline documents (b) protein levels

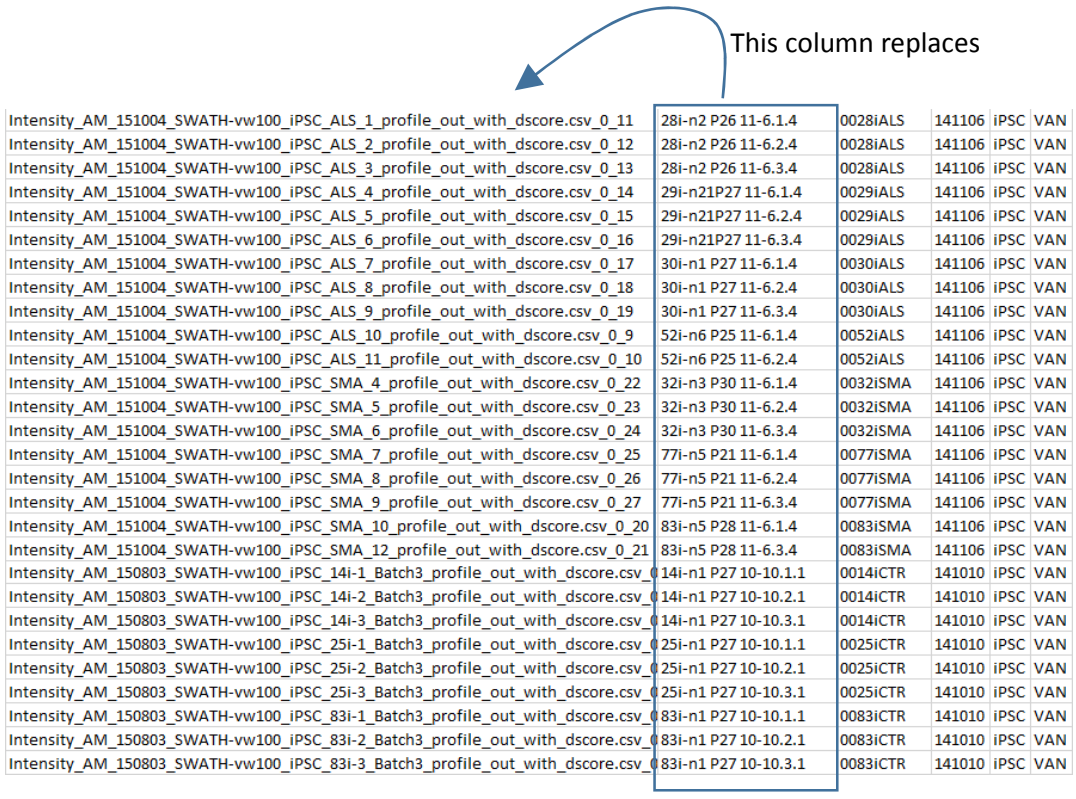
A. Un-normalized peptide data

see file: outputmatrix_OpenSWATH
unformatted tab - IPSC_DDA_6600_QE_combined_Canon
formatted tab - Sept2016_Release

1. File: output matrix generated in OpenSWATH
2. Remove DECOY hits

Peptide	Protein
DECOY_12463_LC(UniMod:4)YVALDFEQEMATVASSSSLEK_3_run0	DECOY_1/sp|A5A3E0|POTEF_HUMAN
DECOY_13837_LNDTLLLGPDPLGNFLSIAVK_3_run0	DECOY_1/sp|O00178|GTPB1_HUMAN
DECOY_15500_MTDLLEEGITVVENIYK_3_run0	DECOY_1/sp|O00186|STXB3_HUMAN
DECOY_13628_LLKEGEEPTVY(UniMod:21)SDEEEPKDESAR_4_run0	DECOY_1/sp|O00264|PGRC1_HUMAN
DECOY_13629_LLKEGEEPT(UniMod:21)VYSDEEEPKDESAR_4_run0	DECOY_1/sp|O00264|PGRC1_HUMAN

3. Replace sample 'intensity_' column headers with official sample generation identifiers:



Intensity_AM_151004_SWATH-vw100_iPSC_ALS_1_profile_out_with_dscore.csv_0_11	28i-n2 P26 11-6.1.4	0028iALS	141106	iPSC	VAN
Intensity_AM_151004_SWATH-vw100_iPSC_ALS_2_profile_out_with_dscore.csv_0_12	28i-n2 P26 11-6.2.4	0028iALS	141106	iPSC	VAN
Intensity_AM_151004_SWATH-vw100_iPSC_ALS_3_profile_out_with_dscore.csv_0_13	28i-n2 P26 11-6.3.4	0028iALS	141106	iPSC	VAN
Intensity_AM_151004_SWATH-vw100_iPSC_ALS_4_profile_out_with_dscore.csv_0_14	29i-n21P27 11-6.1.4	0029iALS	141106	iPSC	VAN
Intensity_AM_151004_SWATH-vw100_iPSC_ALS_5_profile_out_with_dscore.csv_0_15	29i-n21P27 11-6.2.4	0029iALS	141106	iPSC	VAN
Intensity_AM_151004_SWATH-vw100_iPSC_ALS_6_profile_out_with_dscore.csv_0_16	29i-n21P27 11-6.3.4	0029iALS	141106	iPSC	VAN
Intensity_AM_151004_SWATH-vw100_iPSC_ALS_7_profile_out_with_dscore.csv_0_17	30i-n1 P27 11-6.1.4	0030iALS	141106	iPSC	VAN
Intensity_AM_151004_SWATH-vw100_iPSC_ALS_8_profile_out_with_dscore.csv_0_18	30i-n1 P27 11-6.2.4	0030iALS	141106	iPSC	VAN
Intensity_AM_151004_SWATH-vw100_iPSC_ALS_9_profile_out_with_dscore.csv_0_19	30i-n1 P27 11-6.3.4	0030iALS	141106	iPSC	VAN
Intensity_AM_151004_SWATH-vw100_iPSC_ALS_10_profile_out_with_dscore.csv_0_9	52i-n6 P25 11-6.1.4	0052iALS	141106	iPSC	VAN
Intensity_AM_151004_SWATH-vw100_iPSC_ALS_11_profile_out_with_dscore.csv_0_10	52i-n6 P25 11-6.2.4	0052iALS	141106	iPSC	VAN
Intensity_AM_151004_SWATH-vw100_iPSC_SMA_4_profile_out_with_dscore.csv_0_22	32i-n3 P30 11-6.1.4	0032iSMA	141106	iPSC	VAN
Intensity_AM_151004_SWATH-vw100_iPSC_SMA_5_profile_out_with_dscore.csv_0_23	32i-n3 P30 11-6.2.4	0032iSMA	141106	iPSC	VAN
Intensity_AM_151004_SWATH-vw100_iPSC_SMA_6_profile_out_with_dscore.csv_0_24	32i-n3 P30 11-6.3.4	0032iSMA	141106	iPSC	VAN
Intensity_AM_151004_SWATH-vw100_iPSC_SMA_7_profile_out_with_dscore.csv_0_25	77i-n5 P21 11-6.1.4	0077iSMA	141106	iPSC	VAN
Intensity_AM_151004_SWATH-vw100_iPSC_SMA_8_profile_out_with_dscore.csv_0_26	77i-n5 P21 11-6.2.4	0077iSMA	141106	iPSC	VAN
Intensity_AM_151004_SWATH-vw100_iPSC_SMA_9_profile_out_with_dscore.csv_0_27	77i-n5 P21 11-6.3.4	0077iSMA	141106	iPSC	VAN
Intensity_AM_151004_SWATH-vw100_iPSC_SMA_10_profile_out_with_dscore.csv_0_20	83i-n5 P28 11-6.1.4	0083iSMA	141106	iPSC	VAN
Intensity_AM_151004_SWATH-vw100_iPSC_SMA_12_profile_out_with_dscore.csv_0_21	83i-n5 P28 11-6.3.4	0083iSMA	141106	iPSC	VAN
Intensity_AM_150803_SWATH-vw100_iPSC_14i-1_Batch3_profile_out_with_dscore.csv_0	14i-n1 P27 10-10.1.1	0014iCTR	141010	iPSC	VAN
Intensity_AM_150803_SWATH-vw100_iPSC_14i-2_Batch3_profile_out_with_dscore.csv_0	14i-n1 P27 10-10.2.1	0014iCTR	141010	iPSC	VAN
Intensity_AM_150803_SWATH-vw100_iPSC_14i-3_Batch3_profile_out_with_dscore.csv_0	14i-n1 P27 10-10.3.1	0014iCTR	141010	iPSC	VAN
Intensity_AM_150803_SWATH-vw100_iPSC_25i-1_Batch3_profile_out_with_dscore.csv_0	25i-n1 P27 10-10.1.1	0025iCTR	141010	iPSC	VAN
Intensity_AM_150803_SWATH-vw100_iPSC_25i-2_Batch3_profile_out_with_dscore.csv_0	25i-n1 P27 10-10.2.1	0025iCTR	141010	iPSC	VAN
Intensity_AM_150803_SWATH-vw100_iPSC_25i-3_Batch3_profile_out_with_dscore.csv_0	25i-n1 P27 10-10.3.1	0025iCTR	141010	iPSC	VAN
Intensity_AM_150803_SWATH-vw100_iPSC_83i-1_Batch3_profile_out_with_dscore.csv_0	83i-n1 P27 10-10.1.1	0083iCTR	141010	iPSC	VAN
Intensity_AM_150803_SWATH-vw100_iPSC_83i-2_Batch3_profile_out_with_dscore.csv_0	83i-n1 P27 10-10.2.1	0083iCTR	141010	iPSC	VAN
Intensity_AM_150803_SWATH-vw100_iPSC_83i-3_Batch3_profile_out_with_dscore.csv_0	83i-n1 P27 10-10.3.1	0083iCTR	141010	iPSC	VAN

4. Reformat peptide column
    a. Decouple – one column converted to two columns

Peptide		Peptide	Charge (z)
10140_IEDVGSDEEDDSGK_2_run0		IEDVGSDEEDDSGK	2

    b. Replace all UniMod chemical modifications with amu increase.

GADIDALC(UniMod:4)VAPR		GADIDALC(57.0215)VAPR
M(UniMod:35)EGPLSVFGDR		M(15.9949)EGPLSVFGDR

5. Reformat protein column - decouple

Protein		Number of Proteins mapped	UniProtKB	Gene_Organism	UniProtKB	Gene_Organism
2/sp|P08238|HS90B_HUMAN/sp|Q58FF8|H90B2_HUMAN		2	P08238	HS90B_HUMAN	Q58FF8	H90B2_HUMAN
1/sp|P62937|PPIA_HUMAN		1	P62937	PPIA_HUMAN		