

## Appendix

PeakLink: A new peptide peak linking method in LC-MS/MS using wavelet and SVM  
Mehrab Ghanat Bari, Xuepo Ma and Jianqiu Zhang  
Department of Electrical and Computer Engineering  
The University of Texas at San Antonio  
One UTSA Circle  
San Antonio, TX 78246, USA

### 1 Datasets

To developing PeakLink (*PL*) and to test it, we used some freely available datasets, which can be downloaded from [1, 2, 3]. As mentioned in the paper, case 1 to case 7 contain Super-SILAC and case 8 contains label-free datasets.

Table 1 shows the datasets in each case, which aligned with each other. For example, in case 1, *Andromeda* search engine identified 2978 and 2058 peptides for ductal and normal sample respectively, which 1084 peptides are common between two sample. Out of 1084 peptides, we select 468 as training and the rest for testing (616 peptides).

Table 1: Complementary information about the data used in different cases

	Samples name	Fraction/Replicate	Tandem MS features	Intersection	training set
Case 1	20091223_Orbi6_TaGe_QC_ducta_1	1/1	2978	1084	468
	20091223_Orbi6_TaGe_QC_Normal1_01	1/1	2058		
	20091223_Orbi6_TaGe_QC_ducta_2	2/1	4250	1664	718
	20091223_Orbi6_TaGe_QC_Normal1_02	2/1	2975		
	20091223_Orbi6_TaGe_QC_ducta_3	3/1	4167	1594	688
	20091223_Orbi6_TaGe_QC_Normal1_03	3/1	2609		
Case 2	20091223_Orbi6_TaGe_QC_lob_1	1/1	2617	946	408
	20091223_Orbi6_TaGe_QC_Normal1_01	1/1	2058		
	20091223_Orbi6_TaGe_QC_lob_2	2/1	4425	1471	635
	20091223_Orbi6_TaGe_QC_Normal1_02	2/1	2975		
	20091223_Orbi6_TaGe_QC_lob_3	3/1	4257	1325	572
	20091223_Orbi6_TaGe_QC_Normal1_02	3/1	2609		
Case 3	20091223_Orbi6_TaGe_QC_lob_1	1/1	2978	1084	468
	20091223_Orbi6_TaGe_QC_ducta_1	1/1	2978		
Case 4	20091223_Orbi6_TaGe_QC_Normal1_02	2/1	2978	1084	468
	20090608_Orbi6_TaGe_SA_TUMOR_5mix1_02	2/1	2978		
Case 5	20091223_Orbi6_TaGe_QC_ducta_1	1/1	4425	2045	883
	20090608_Orbi6_TaGe_SA_TUMOR_5mix1_01	1/1	5035		
Case 6	20091223_Orbi6_TaGe_QC_ducta_1	1/1	2978	1084	468
	20091230_Orbi6_TaGe_SA_duct3_1	3/1	2978		
Case 7	200090815_Velos5_TaGe_SA_Silacmix_TOP15_01	1/1	4438	1043	450
	20091223_Orbi6_TaGe_QC_lob_1	1/1	2938		
Case 8	20110922_EXQ4_NaNa_SA_YeastEasy_Labelfree_01	–	20400	374	37 – 299
	022008_F4_DTree_1	–	1338		

## References

- [1] Askenazi, M. *et al.* (2011) iprg 2011: A study on the identification of electron transfer dissociation mass spectra. *Journal of Biomolecular Techniques: JBT*, **22**(Supplement), S20.
- [2] Geiger, T. *et al.* (2010) Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nature Methods*, **7**, 383-385.
- [3] Nagaraj, N. *et al.* (2012) System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra hplc runs on a bench top orbitrap. *Molecular and Cellular Proteomics*, **11**(3).