Protein samples

We used six protein samples, digested with trypsin, labeled with ^{18}O-water, and analyzed using high power liquid chromatography (HPLC) and mass spectrometry (MS) (Figure 1). The first sample (2 mg of protein) was prepared from mouse renal cortical extract[1]. The extract was divided into two parts and one was labeled using trypsin digest. Labeled and unlabeled peptides were mixed in 1:1 ratio and analyzed in four runs. The other five samples were prepared from bovine serum albumin (BSA)[2]. Labeled and unlabeled peptides were mixed in ratios of 5:1, 3:1, 1:1, 1:3, and 1:5. Each mixture contained about 2 pmol of BSA tryptic peptides and was analyzed in three runs on a HPLC-MS/MS system.

Liquid chromatography and tandem mass spectrometry

LC-MS/MS experiments were performed on a LTQ linear ion trap mass spectrometer (ThermoFinnigan, San Jose, CA) equipped with a nanospray source; the mass spectrometer was coupled on-line to a ProteomX® nano-HPLC system (ThermoFinnigan, San Jose, CA). Two µL of each peptide solution were manually injected and separated on a reversed-phase nano-HPLC column. The mass spectrometer was operated in the data-dependent triple-play mode. The three most intense ions in each MS survey scan were automatically selected for zoom scan and MS/MS.

Database searching

Tandem mass spectral data were generated from LTQ raw files using Bioworks 3.2. The group scan was set equal to zero, the minimum group count was 1, the ion threshold was 10, the intensity threshold was 100, and the precursor mass range was between 400 and 6000 Da. Database searches were conducted using TurboSEQUEST[3] v. 27 (rev. 12). The precursor ion mass tolerance was set at 3.0 Da and the fragment ion tolerance was set at 1.0 Da. Methionine residues were searched as being differentially modified (16.0 Da) due to oxidation. The C-termini of peptides were differentially modified by 4.0 Da. Cysteine residues were modified (57.0215 Da) statically to account for alkylation with iodoacetamide. Trypsin was specified for enzymatic cleavage, and up to 3 missed cleavage sites were allowed. The mouse dataset was searched against the database containing mouse proteins (downloaded from SwissProt website on March 13, 2008) combined with reversed sequences (30,722 entries altogether). The BSA dataset was searched against a database of 15,988 proteins that contained the BSA and yeast protein sequences (downloaded from the *Saccharomyces* Genome Database on August 21, 2007), and their reversed complements. We used a deltaCn cut-off value of 0.1 to filter peptide sequence matches; the false discovery rate (FDR) for peptides was set at 3%.

The spectral datasets were converted into mzML and pep.XML file formats using msconvert and Out2XML programs in the Trans Proteomics Pipeline[4;5] suit of programs. MassXplorer was used to quantify peptides.

Peptide ratio estimations

If we denote the abundances of the unlabeled species as A, singly ^{18}O labeled species as B_1, and doubly ^{18}O labeled species as B_2, then they are related to the experimentally observed ion abundance levels via the following relationships[6]:

$$I_0 = A$$
$$I_2 = \frac{M_2}{M_0} * A + B_1$$
$$I_4 = \frac{M_4}{M_0} * A + \frac{M_2}{M_0} B_1 + B_2$$

where I_0, I_2 and I_4 are the abundance levels of the monoisotopic, second (overlapping with the monoisotopic peak of the singly ^{18}O labeled peptide) and fourth (overlapping with the monoisotopic peak of the doubly ^{18}O labeled peptide) isotopic peaks; and M_0, M_2 and M_4 are the theoretical relative abundance levels of the monoisotopic, second and fourth naturally occurring isotopic peaks, respectively. The ratio of peptide pairs, R, is determined from the solutions of the above equations for A, B_1 and B_2:

$$R = \frac{B_1 + B_2}{A} = \frac{I_4 + (1 - M_2 / M_0) I_2 - (M_4 / M_0) I_0 - (1 - M_2 / M_0)(M_2 / M_0) I_0}{I_o}$$

The above formulae assume that the portion of the second isotopic peak not accounted for by the monoisotopic peak of the unlabeled peptide is due to a single ^{18}O-labeled peptide.

Reference List

1. Zhao,Y., Denner,L., Haidacher,S.J., LeJeune,W.S., and Tilton,R.G. (2008) Comprehensive analysis of the mouse renal cortex using two-dimensional HPLC – tandem mass spectrometry. Proteome Science, 6,15.

2. Qian,W.J., Monroe,M.E., Liu,T., Jacobs,J.M., Anderson,G.A., Shen,Y., Moore,R.J., Anderson,D.J., Zhang,R., Calvano,S.E., Lowry,S.F., Xiao,W., Moldawer,L.L., Davis,R.W., Tompkins,R.G., Camp,D.G., and Smith,R.D. (2005) Quantitative proteome analysis of human plasma following in vivo lipopolysaccharide administration using 16O/18O labeling and the accurate mass and time tag approach. Mol.Cell Proteomics, 4,700-709.

3. Eng,J.K., McCormack,A.L., and Yates,J.R., III (1994) An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. J.Am.Soc.Mass Spectrom., 5,976-989.

4. Keller,A., Eng,J., Zhang,N., Li,X.J., and Aebersold,R. (2005) A uniform proteomics MS/MS analysis platform utilizing open XML file formats. Mol.Syst.Biol., 1,2005.

5. Deutsch,E.W., Mendoza,L., Shteynberg,D., Farrah,T., Lam,H., Tasman,N., Sun,Z., Nilsson,E., Pratt,B., Prazen,B., Eng,J.K., Martin,D.B., Nesvizhskii,A.I., and Aebersold,R. (2010) A guided tour of the Trans-Proteomic Pipeline. Proteomics., 10,1150-1159.

6. Yao,X., Afonso,C., and Fenselau,C. (2003) Dissection of proteolytic 18O labeling: endoprotease-catalyzed 16O-to-18O exchange of truncated peptide substrates. J.Proteome.Res., 2,147-152.

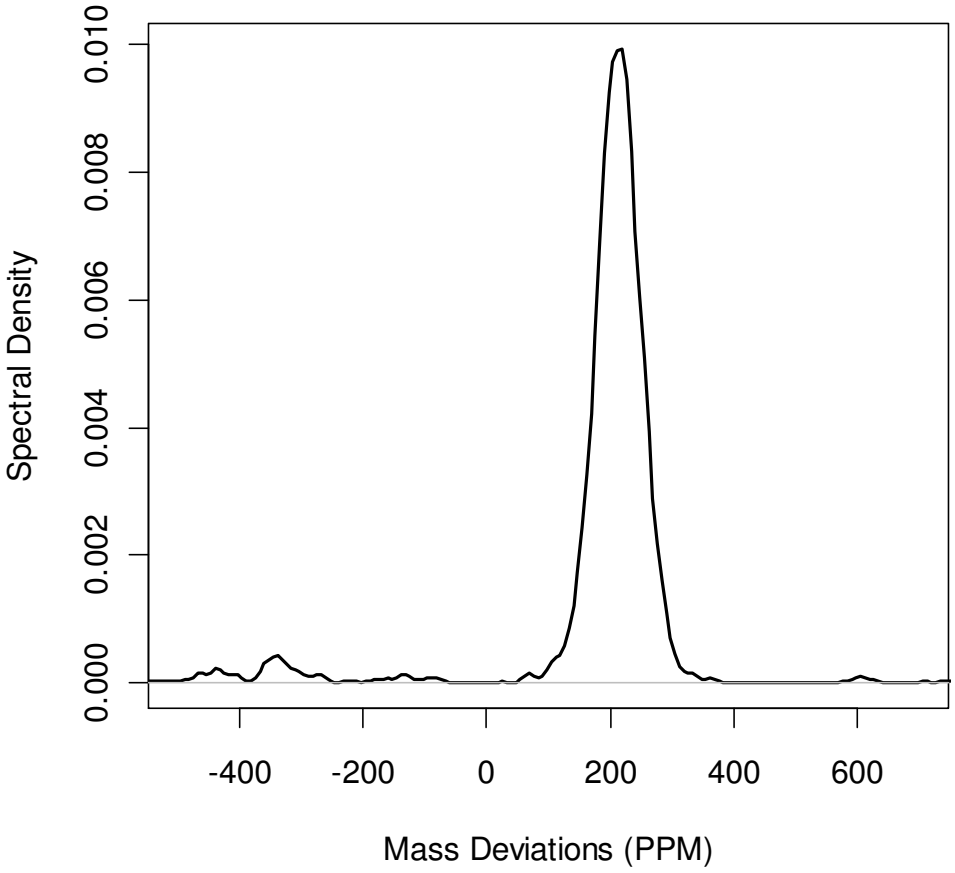

Figure S1. Density of mass deviations of monoisotopic peak positions for light peptides from the experimentally measured values using the mouse data set. In this sample, heavy and light peptides were mixed in 1:1 ratio. The distribution has a global maximum of 200 PPM. While the mode value may be different for other datasets, the distribution is always centered and the scaled value of the mass deviations are used as features for classifying mass profiles. In this training dataset, the average peptide ratio in the 200 PPM interval of the median is 1.2. The variance of the ratios in this interval is 0.1 (1042 spectra). The mean and variance of the spectra outside of the 200 PPM interval from the median mass difference were 1.8 and 12.2. These values are in contrast to the unfiltered data, which has a mean of 1.3 and a variance of 2.2 (2003 spectra).

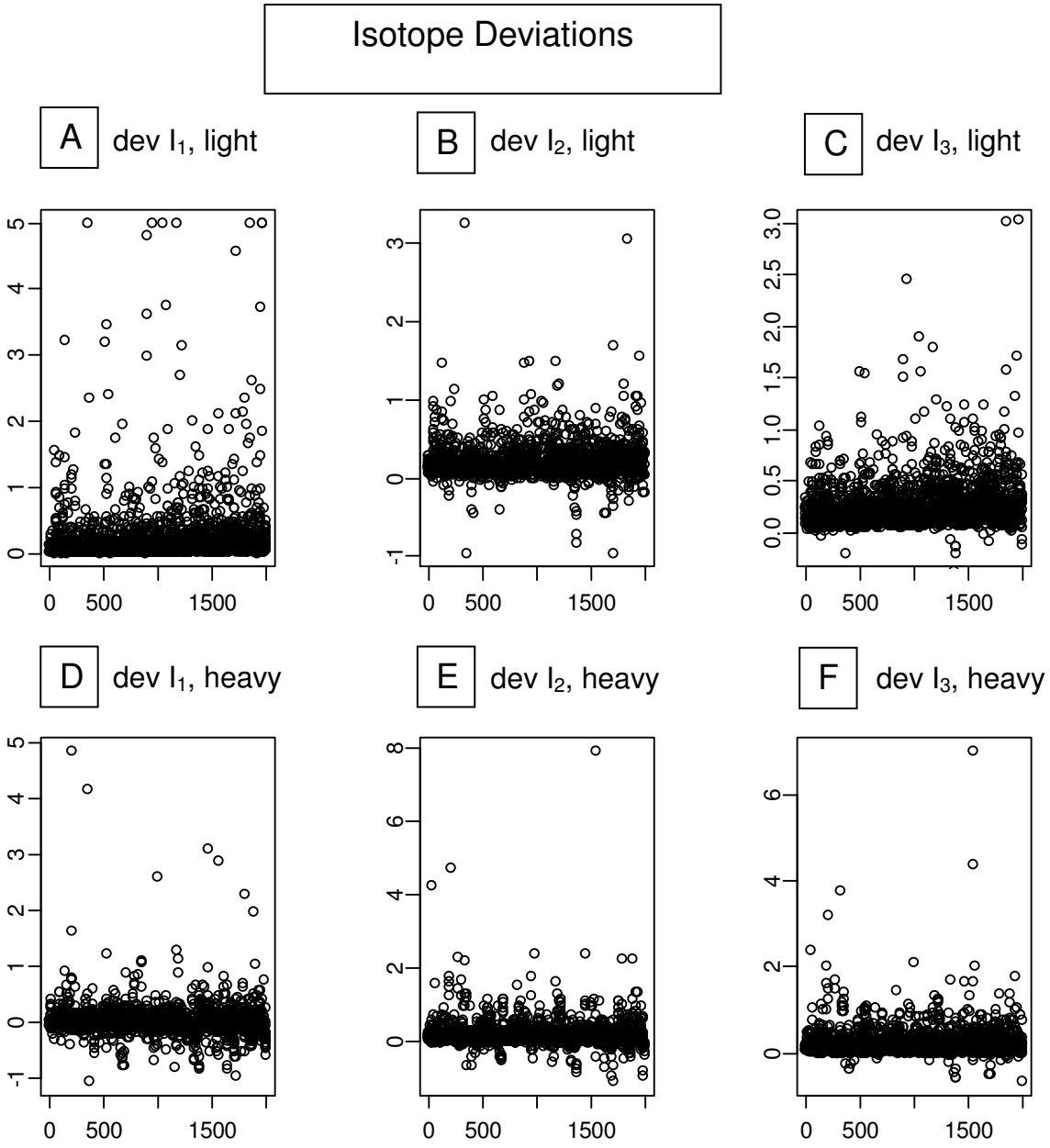

Figure S2. Deviations of isotope abundances from the theoretically expected. In each plot, the horizontal axis shows scan numbers. Plots A-C show deviations for light peptides; D-F show deviations for heavy peptides. All distributions are structured. The distributions were obtained from the mouse sample (1:1 ratio of heavy and light peptides).

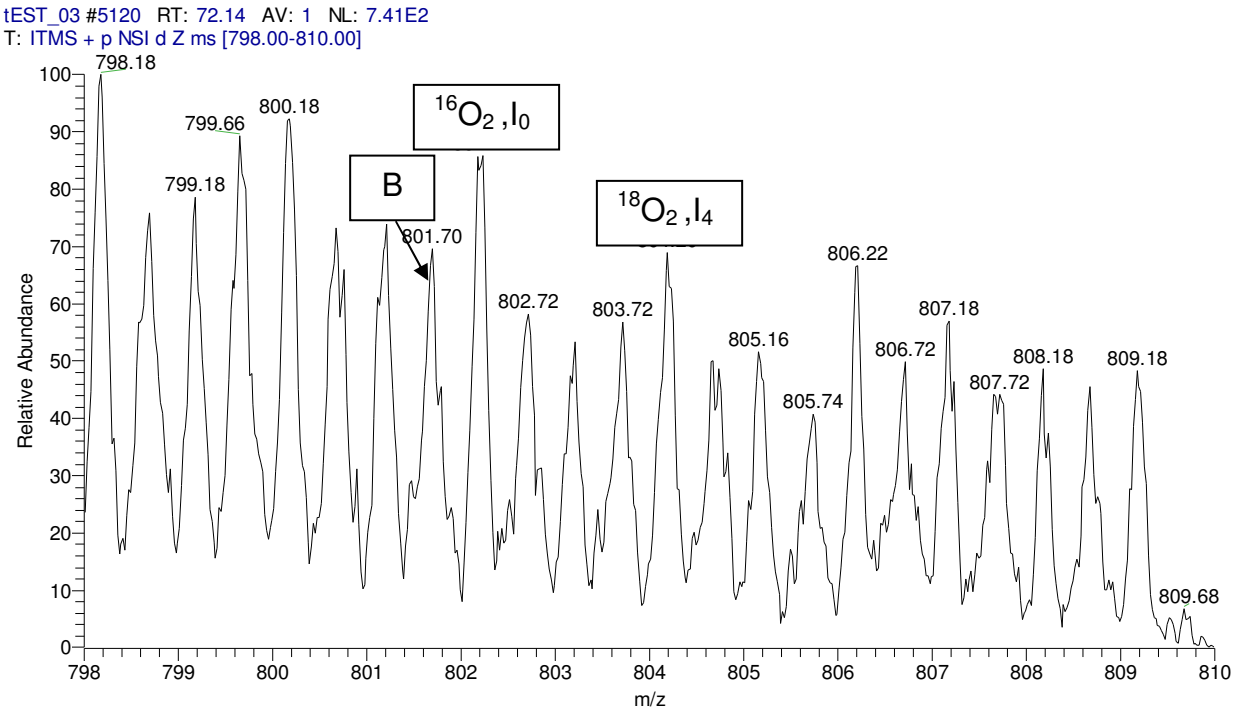

Figure S3. Effect of the preceding peak on ratio estimation. I_0 is the monoisotopic peak of the light peptide, B is the preceding peak, and I_4 is the peak of the heavy peptide. If the preceding peak B is comparable in abundance to the monoisotopic peak I_0, then ratio estimations may be unreliable.

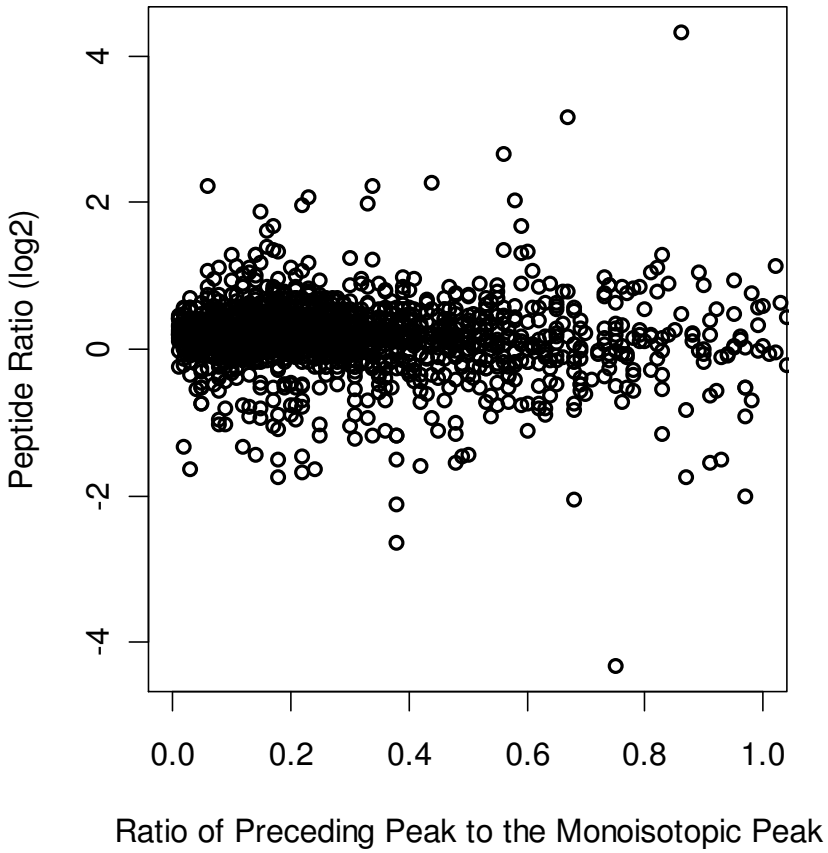

Figure S4. Distribution of the peptide ratio logarithms (2-based) versus ratios of the preceding peak to the monoisotopic peak. For the large values of the peak ratio, the heavy to light peptide ratios show larger variances.

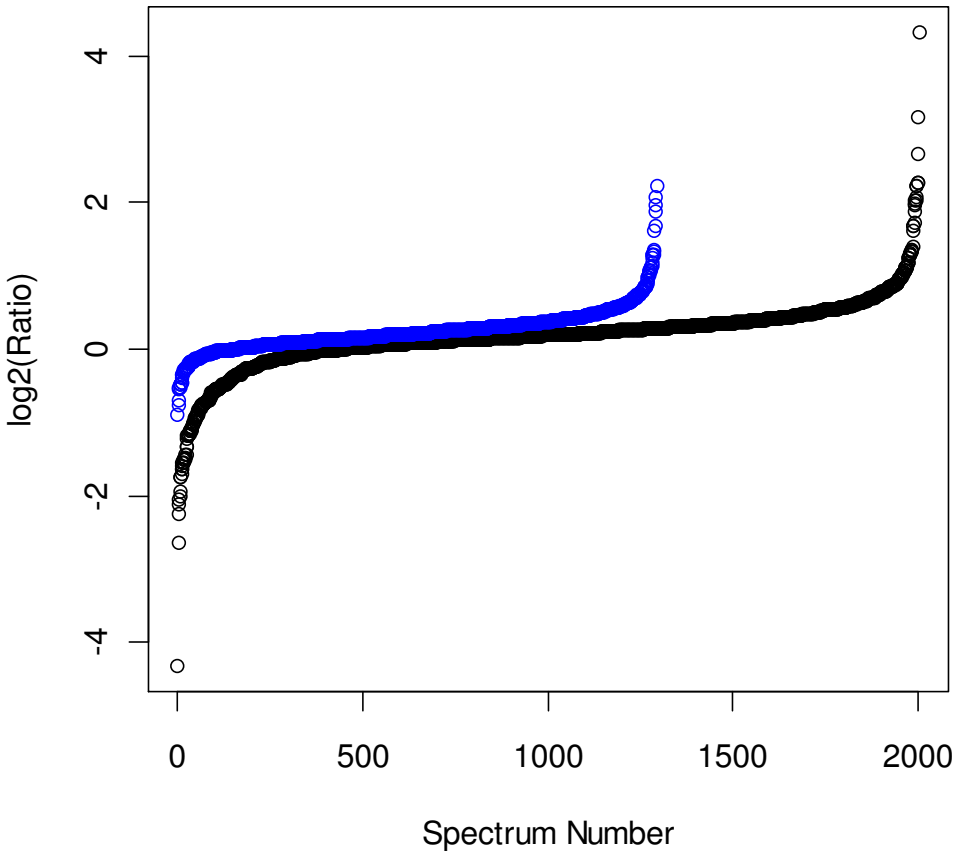

Figure S5. Distribution of ratios before (black) and after (blue) filtering by SVM. The expected ratio in this sample is 1:1. SVM filtering significantly reduces the errors associated with ratios. The variance of the ratios is reduced by five times, from 0.4 to 0.08.

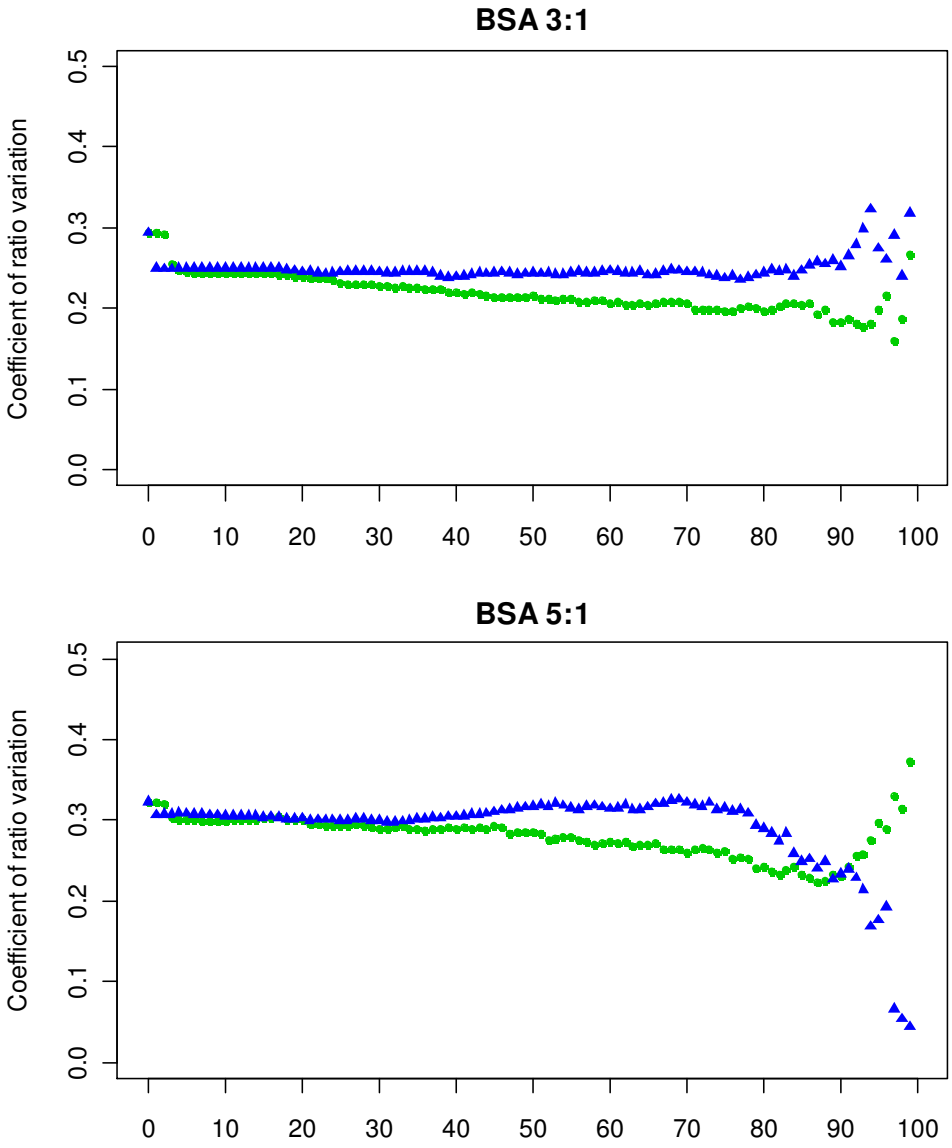

Figure S6. Coefficient of ratio variation versus percentage of spectra filtered by SVM (green circles) and S/N (blue triangles) for two BSA samples mixed in 3(H):1(L) and 5(H):1(L) ratios. For these datasets, the original, unfiltered CV's were relatively small (about 0.3), and their improvement obtained by both filtering methods was not as dramatic as for other samples.