Supplemental Figure 5: For all proteins covered by the PeptideAtlas (PA) repository (October 2017) the most N-terminal peptides were extracted if they had semi-trypsin or Arg-C specificity (C-terminal specific, N-terminal flexible), corresponding to 11,282 PA trypsin and 7,868 PA Arg-C peptides.

A comparison of peptide lengths (in number of amino acids) shows a slightly more narrow distribution for our dataset. Our dataset and the PA trypsin peptides have a median length of 13 amino acids (dashed lines), whereas PA Arg-C peptides have a median of 14 amino acids. Since peptide length slightly correlates with net-charge (more basic amino acids), long highly charged Arg-C peptides are underrepresented in our dataset, in accordance with the iTRAQ-related issues mentioned in Supplemental Figure 4.