

Relative quality of reference versus SAP peptide PEP scores

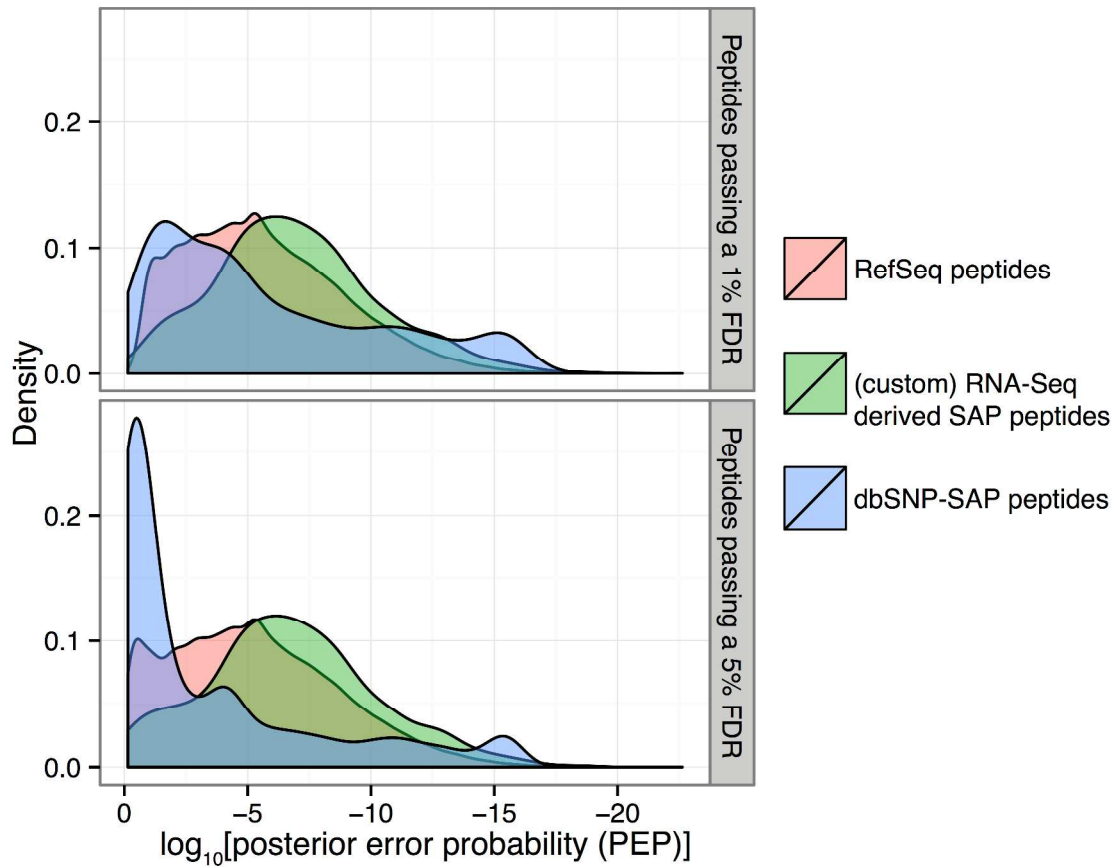


Figure S1: Comparing the posterior probability score (PEP) distributions between RefSeq, (custom) SAP, and dbSNP-SAP peptides. The PEP value is the probability that the peptide match is incorrect¹. The x-axis is plotted to show better matching peptides to the right. The PEP values between peptide groups analyzed show a similar trend to the XCorr comparison.

Relative quality of reference versus SAP peptide q-values

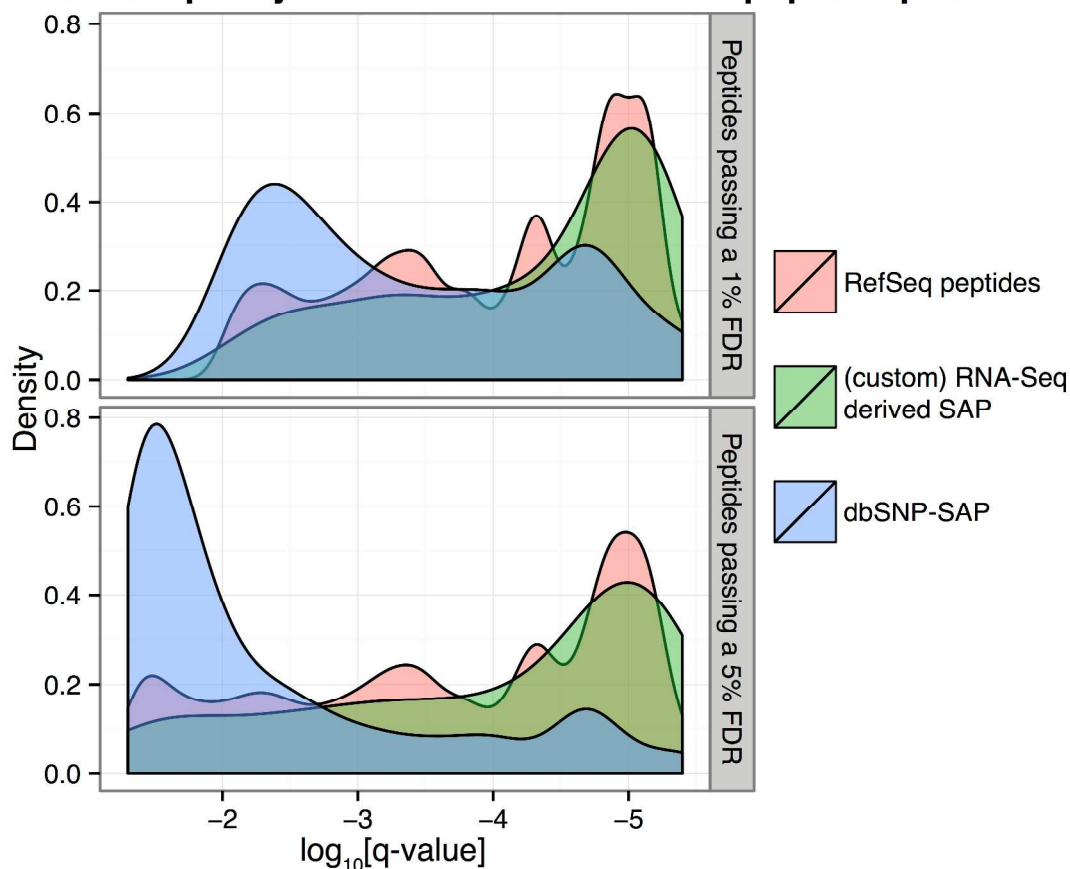


Figure S2: Comparing the q-value distributions between RefSeq, (custom) SAP, and dbSNP-SAP peptides. For a given peptide spectral match (PSM), the q-value represents the minimum FDR threshold at which the peptide will still pass the FDR cut-off¹. The x-axis is plotted to show better matching peptides to the right. The q-values between peptide groups analyzed show a similar trend to the XCorr comparison.

Comparison of XCorr score distributions*: dbSNP-SAP peptides versus RefSeq peptides

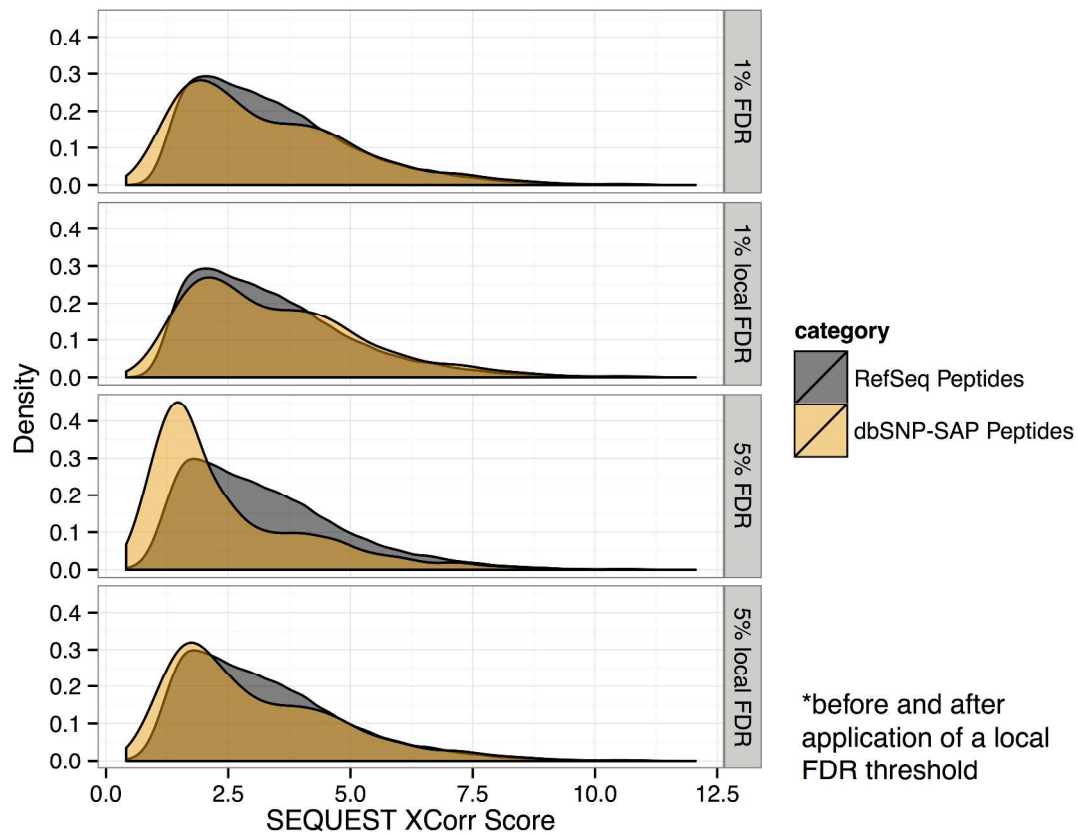


Figure S3: Comparison of dbSNP-SAP and RefSeq peptide XCorr distributions before and after application of a local FDR cut-off for dbSNP-SAP peptides. The local FDR was calculated by dividing the sum of posterior error probability (PEP) values by the number of peptides in the group.

XCORR Distribution of custom (RNA-Seq derived) SAP vs RefSeq Peptides

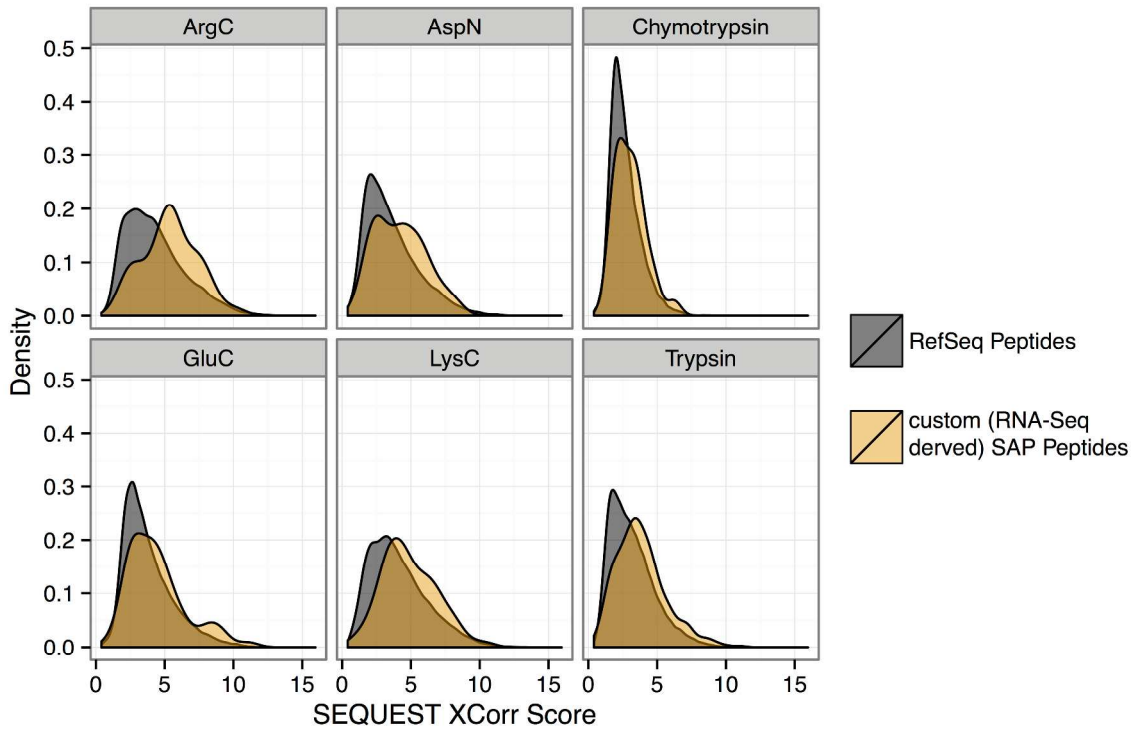


Figure S4: Comparison of custom (RNA-Seq derived) SAP and RefSeq peptide XCORR distributions for each of six protease digest results. For all six peptide populations, the SAP peptides had, on average, slightly better XCORRs than the RefSeq (canonical) peptide identifications.

Transcript ~ Protein Abundance Correlation

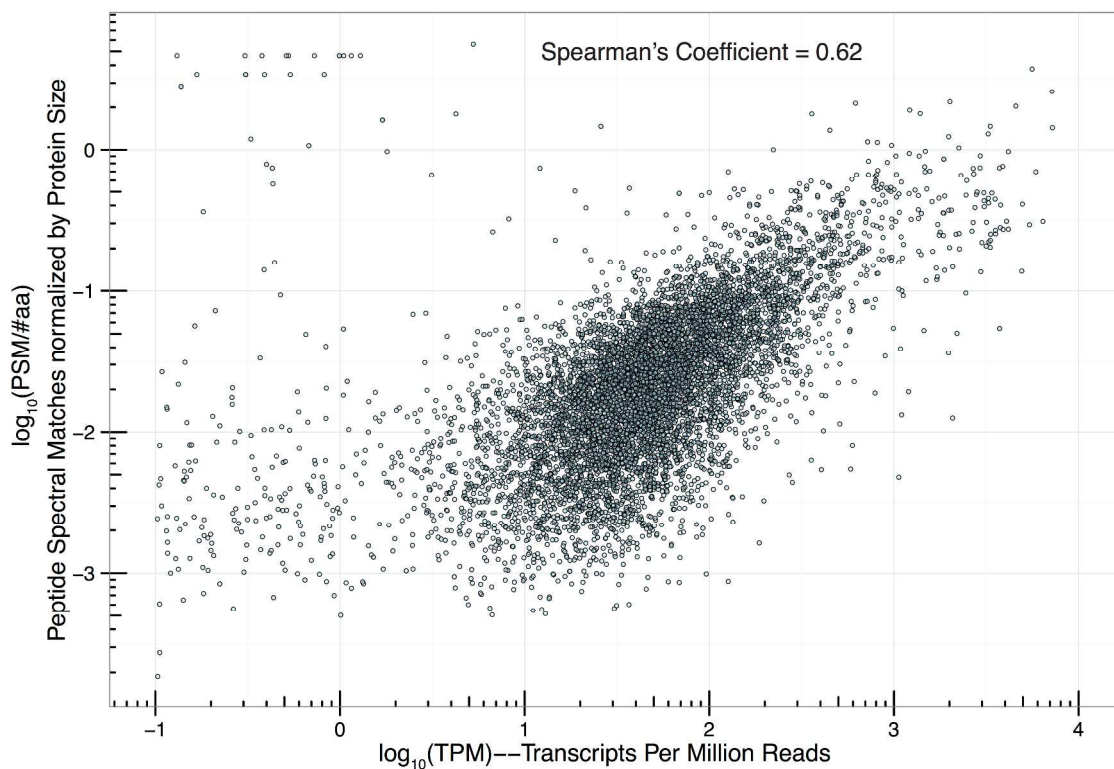


Figure S5: A plot of the relative transcript versus protein abundance levels for each detected gene in the Jurkat cell line (represents 7354 unique NP RefSeq protein entries). TPM values, or Transcripts Per Million of RNA-Seq reads, were generated using RSEM software². PSM values, or peptide spectral matches, were normalized by dividing by the number of amino acids in the corresponding protein. The Spearman's rank correlation coefficient was 0.62.

REFERENCES:

1. Käll, L.; Storey, J. D.; MacCoss, M. J.; Noble, W. S., Posterior Error Probabilities and False Discovery Rates: Two Sides of the Same Coin. *Journal of Proteome Research* **2007**, 7, (1), 40-44.
2. Li, B.; Dewey, C. N., RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *Bmc Bioinformatics* **2011**, 12, 323.