

## **Supporting Information**

### **Screening a Resource of Recombinant Protein Fragments for Targeted Proteomics**

Fredrik Edfors<sup>1</sup>, Björn Forsström<sup>1</sup>, Helian Vunk<sup>1</sup>, David Kotol<sup>1</sup>, Claudia Fredolini<sup>2</sup>, Gianluca Maddalo<sup>1</sup>, Anne-Sophie Svensson<sup>3</sup>, Tove Boström<sup>3,4</sup>, Hanna Tegel<sup>3</sup>, Peter Nilsson<sup>2</sup>, Jochen M. Schwenk<sup>2</sup> and Mathias Uhlen<sup>1,3,5,6</sup>

<sup>1</sup>Science for Life Laboratory, Division of Systems Biology, Department of Protein Science, KTH - Royal Institute of Technology, SE-171 21 Stockholm, Sweden

<sup>2</sup>Science for Life Laboratory, Division of Affinity Proteomics, Department of Protein Science, KTH - Royal Institute of Technology, SE-171 21 Stockholm, Sweden

<sup>3</sup>Albanova University Center, KTH – Royal Institute of Technology, SE – 171 21 Stockholm, Sweden

<sup>4</sup>Atlas Antibodies AB, SE- 114 21 Stockholm, Sweden

<sup>5</sup>Department of Neuroscience - Karolinska Institute, SE - 171 65 Solna, Sweden

<sup>6</sup>Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, DK-2970 Hørsholm, Denmark

# Table of Contents

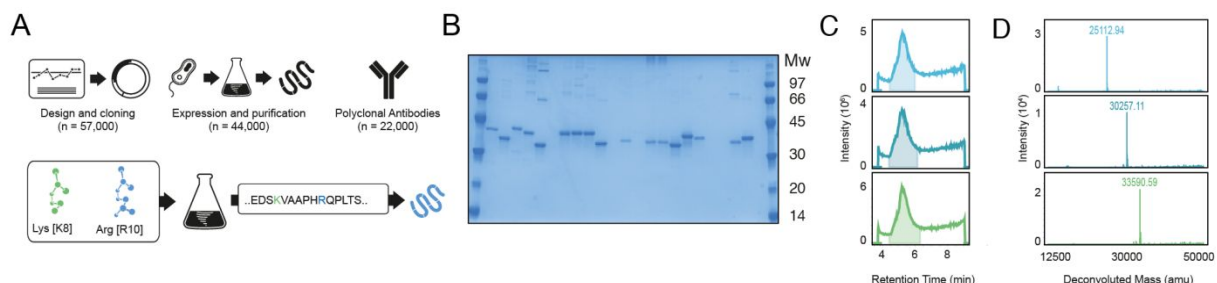
## Supplementary Figures

- Figure S-1 Overview of the recombinant proteins from the HPA resource.
- Figure S-2 Peptide coverage of the human proteome defined by HPA resource (v18, n = 19,613 genes)
- Figure S-3 Length distribution of PrEST fragments subjected to analysis.
- Figure S-4 Database search using a custom *E. coli* database for each pool respectively
- Figure S-5 All peptides successfully quantified from the digestion kinetics experiment.
- Figure S-6 Proteins quantified by multiple proteotypic peptides.
- Figure S-7 Peptide correlation (Pearson's *r*) across multiple samples for APOA4 peptides (n = 3).

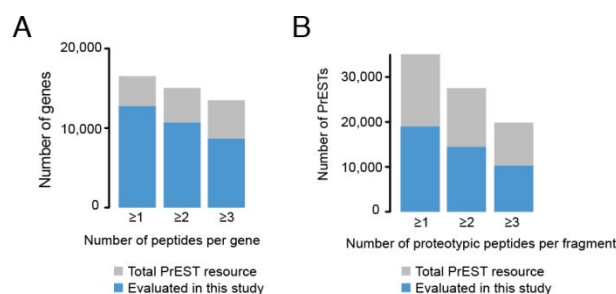
## Supplementary Tables

- Table S-1 PrEST cultivation cost calculation.
- Table S-2 PrEST fragments evaluated in this study
- Table S-3 All theoretical PrEST peptides
- Table S-4 Spike-in amount of SIS PrESTs used for digestion kinetics
- Table S-5 SWATH variable windows
- Table S-6 SIS PrESTs used for PRM quantification
- Table S-7 PRM method details
- Table S-8 Experimental PrEST peptides
- Table S-9 Human peptides with modifications
- Table S-10 Experimental missed cleaved peptides
- Table S-11 iRT for QTag peptides

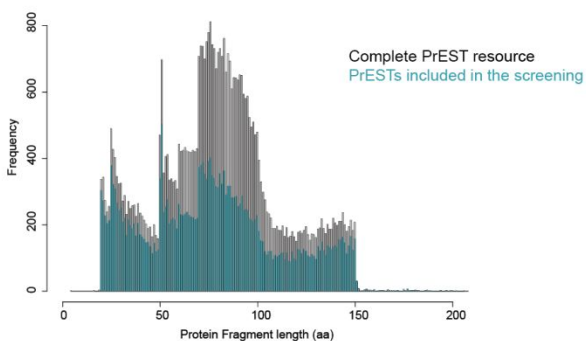
Table S-12	Peptide ratio from digestion kinetics.
Table S-13	Fragment ion intensity for full-length protein digests
Table S-14	Results from absolute quantification of QPrESTs
Table S-15	Intra-individual variation over one year
Table S-16	Inter-individual variation per gene
Table S-17	Experimental peptides per gene and chromosome
Table S-18	Molecular weight for proteins quantified by PRM



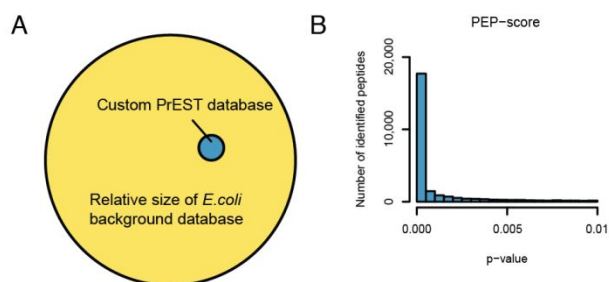
**Figure S-1. Overview of the recombinant protein resource originating from the HPA resource.** **A.** Schematic outline of the PrEST resource consisting of more than 57,000 vectors and 46,700 recombinant protein fragments used for antibody generation. Each PrEST vector can be transformed into an auxotrophic *E. coli* strain to produce a heavy labeled standard. **B.** A representative SDS-PAGE gel of PrEST fragments after individual IMAC purification **C.** Each PrEST fragment is analyzed by LC-MS in to verify the correct intact molecular weight.



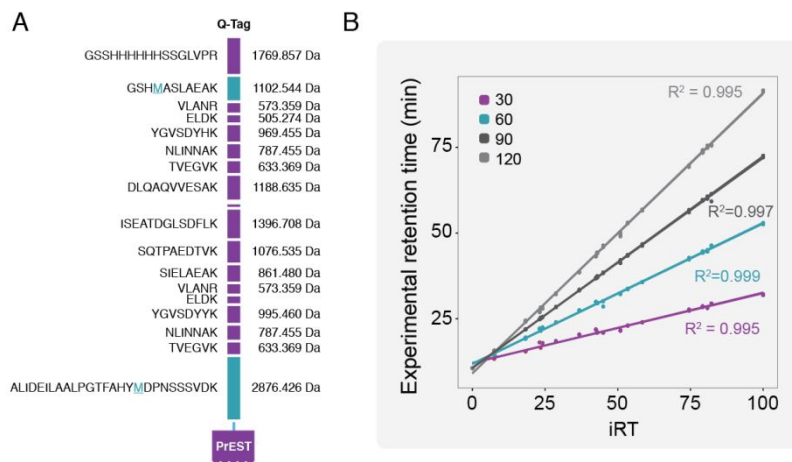
**Figure S-2- Peptide coverage of the human proteome defined by the HPA resource v18 (n = 19,613 genes).** Blue bars represent recombinant proteins screened in this study (n = 26,840) and gray bars correspond to the remaining non-screened resource of recombinant proteins stored in individual microcentrifuge tubes. **A.** Number of genes covered by the HPA resource with at least one proteotypic peptide. **B.** The number of proteotypic peptides per PrEST fragment.



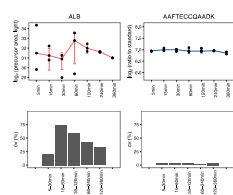
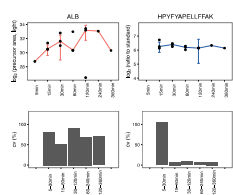
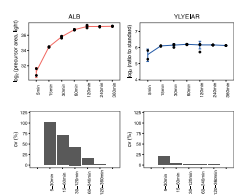
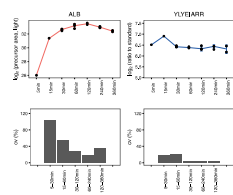
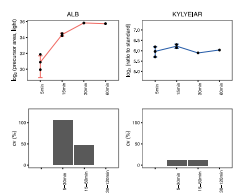
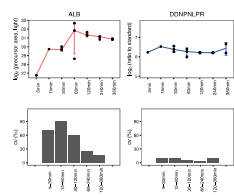
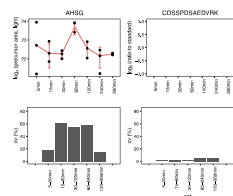
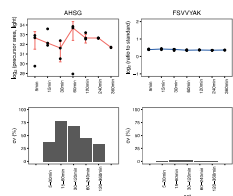
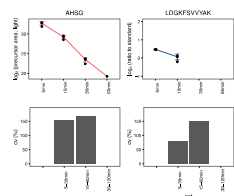
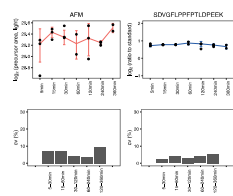
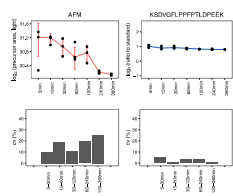
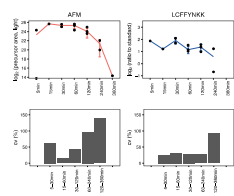
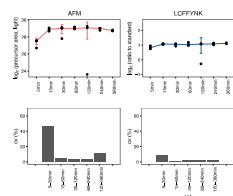
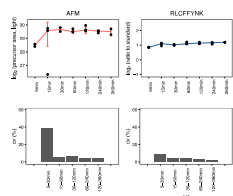
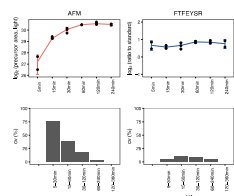
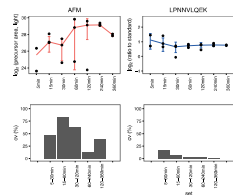
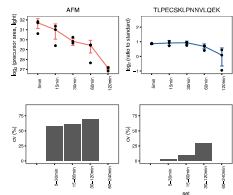
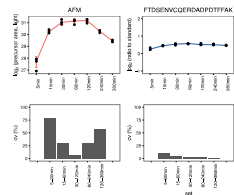
**Figure S-3 –Distribution of number of amino acids per PrEST fragments.** Histogram illustrating the length of all PrEST fragments subjected for LC-MS/MS analysis (blue) in this study (n = 26,840) in comparison to all protein fragments (n = 46,790) available within the complete Human Protein Atlas resource (gray).

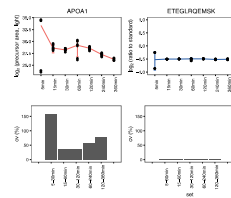
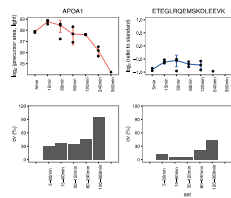
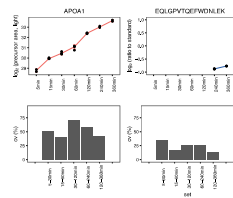
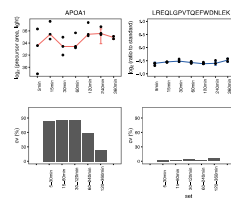
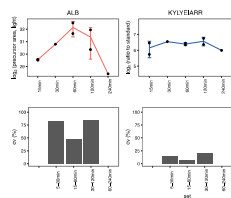
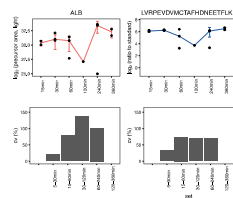
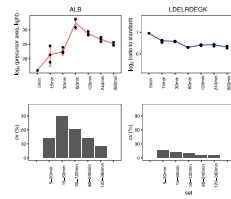
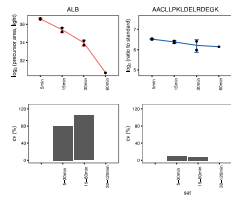
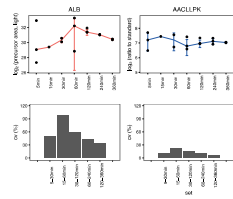
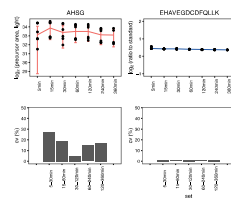
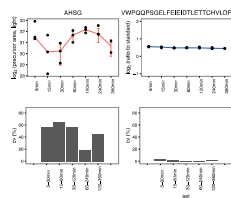
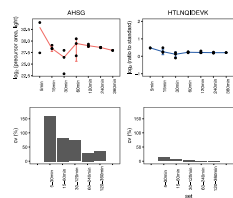
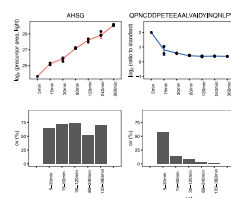
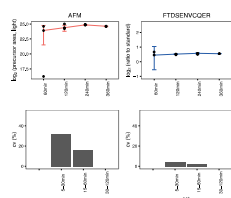
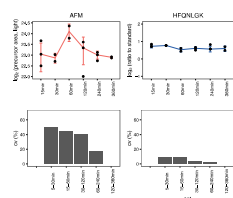
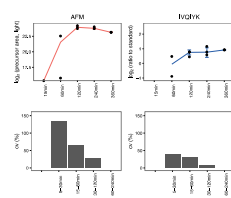
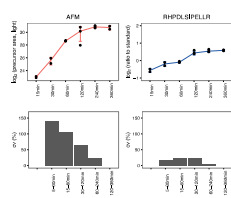
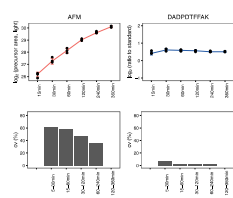


**Figure S-4 – Database search using a custom *E. coli* database for each pool respectively** **A.** Relative size of PrEST library in blue compared to *E. coli* (BL21) in yellow. **B.** Distribution of PEP-scores observed for all identified PrEST peptides included in the final database.

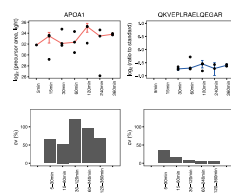
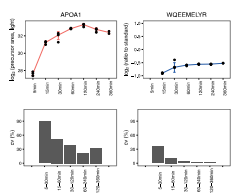
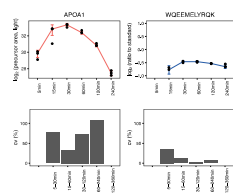
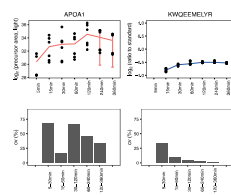
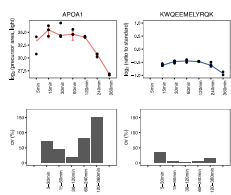
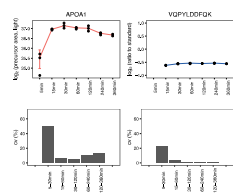
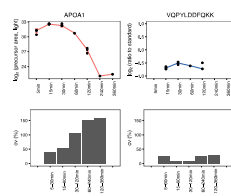
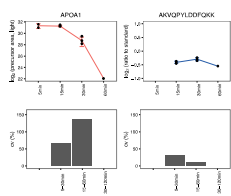
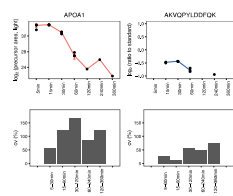
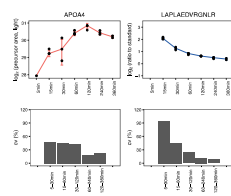
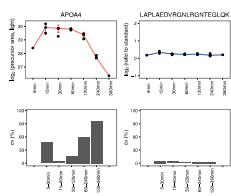
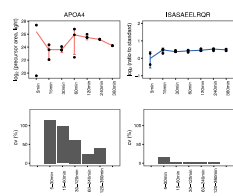
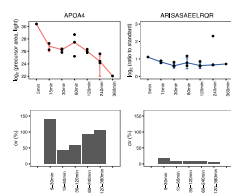
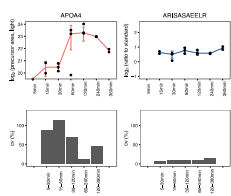
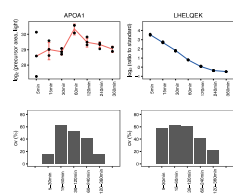
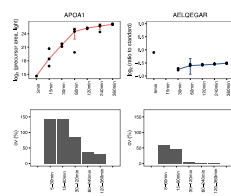
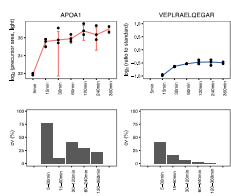
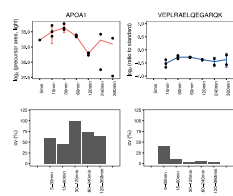


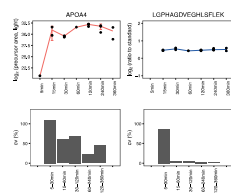
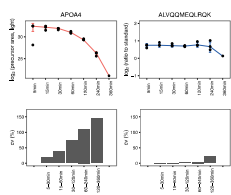
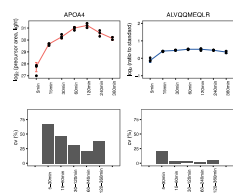
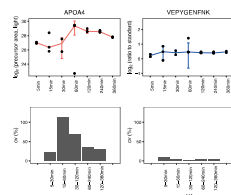
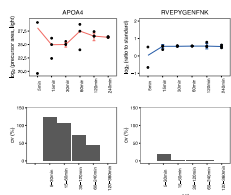
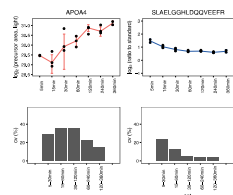
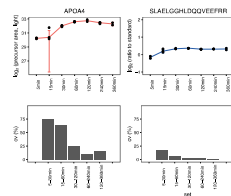
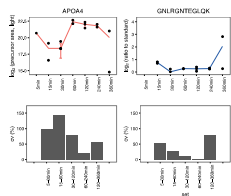
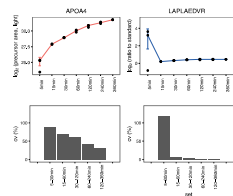
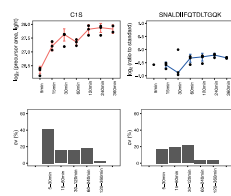
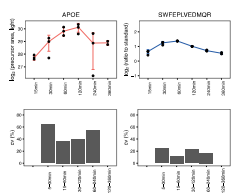
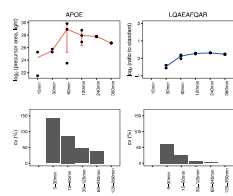
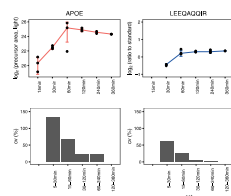
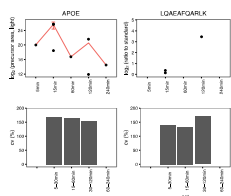
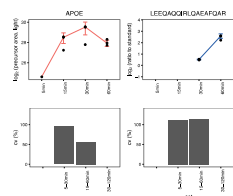
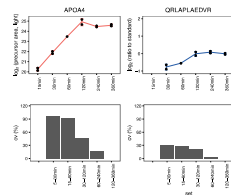
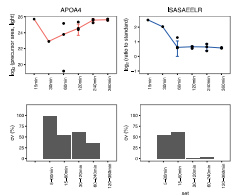
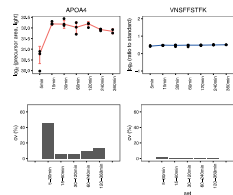
**Figure S-5 Q-tag peptides for retention time calibration.** **A.** Overview of the Q-Tag sequence used for both stable isotope standard quantification and retention time calibration. **B.** Standard curves subjected to retention time alignment based on Q-Tag peptides from digested SIS PrESTs separated on four different linear gradients.

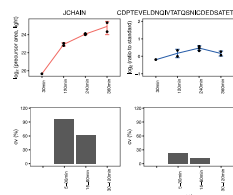
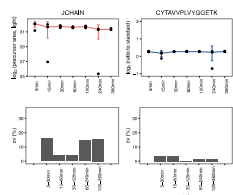
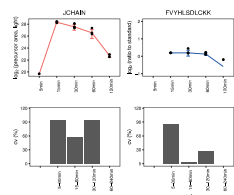
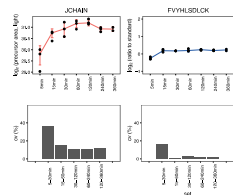
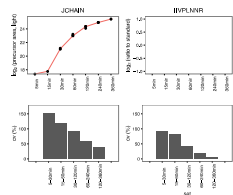
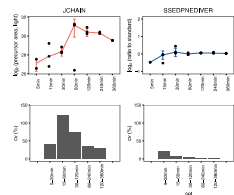
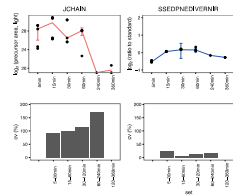
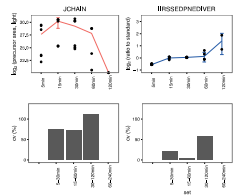
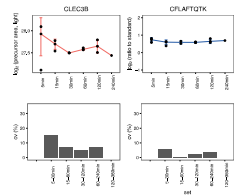
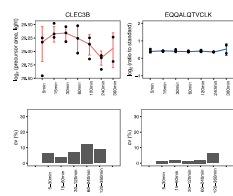
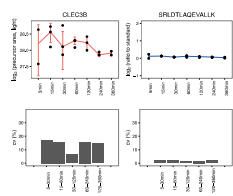
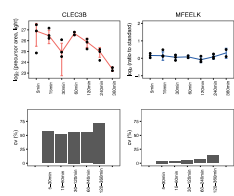
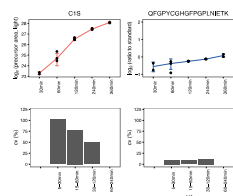
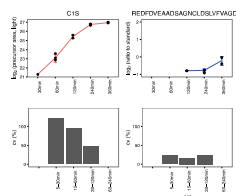
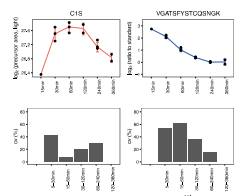
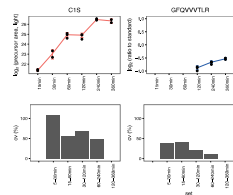
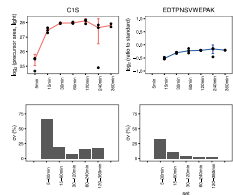
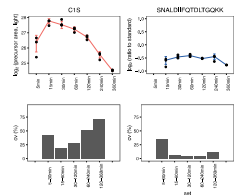


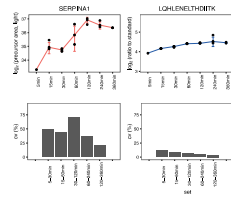
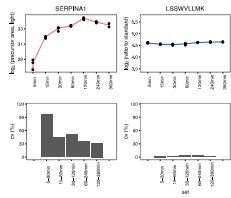
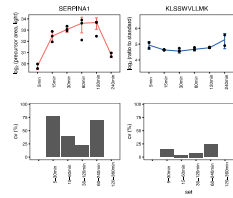
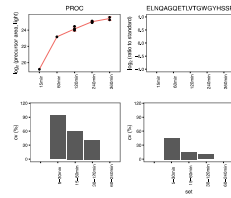
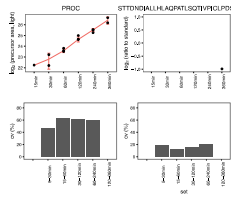
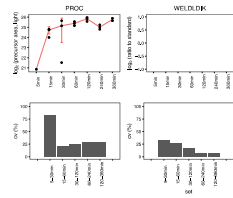
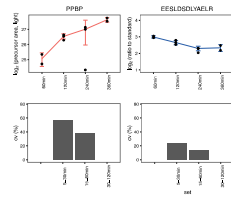
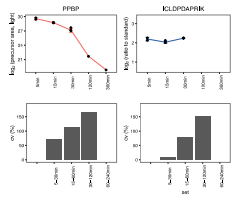
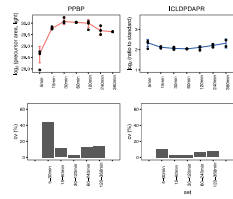
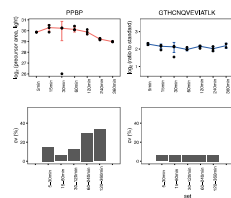
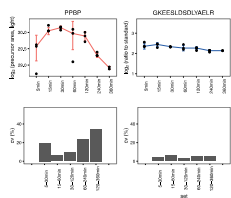
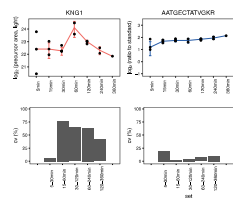
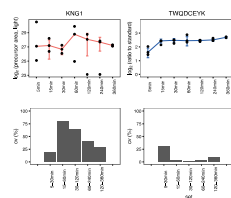
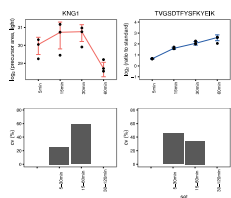
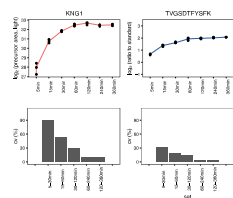
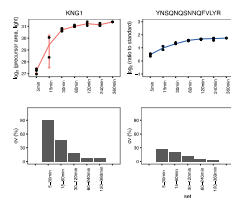
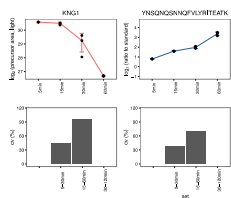
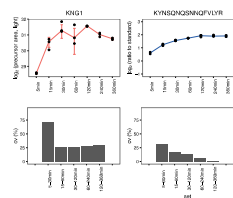


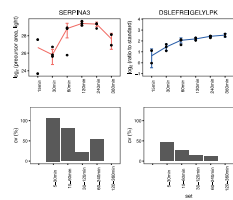
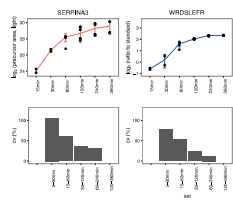
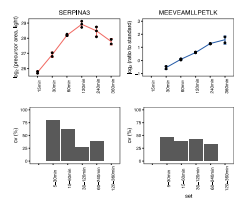
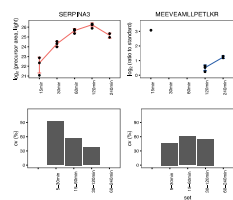
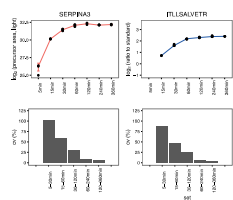
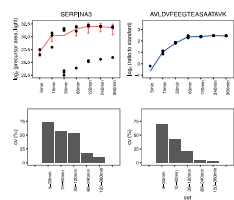
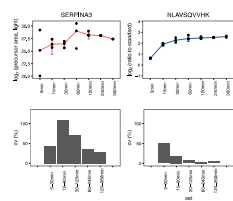
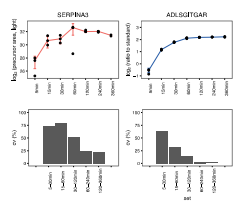
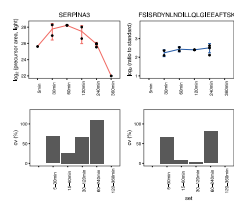
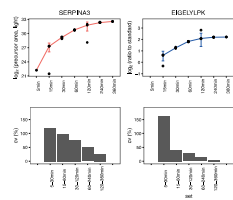
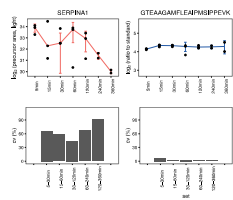
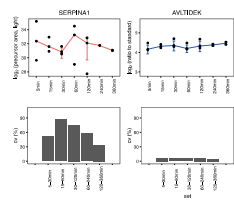
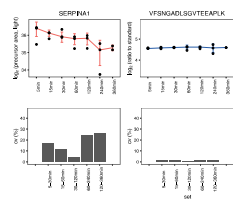
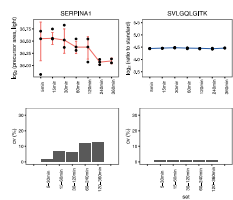
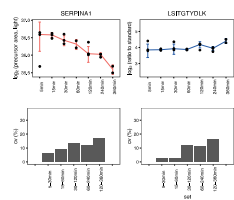
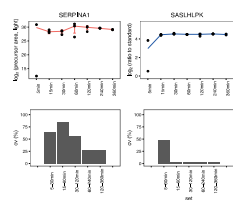
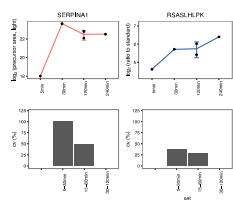
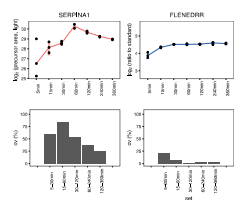


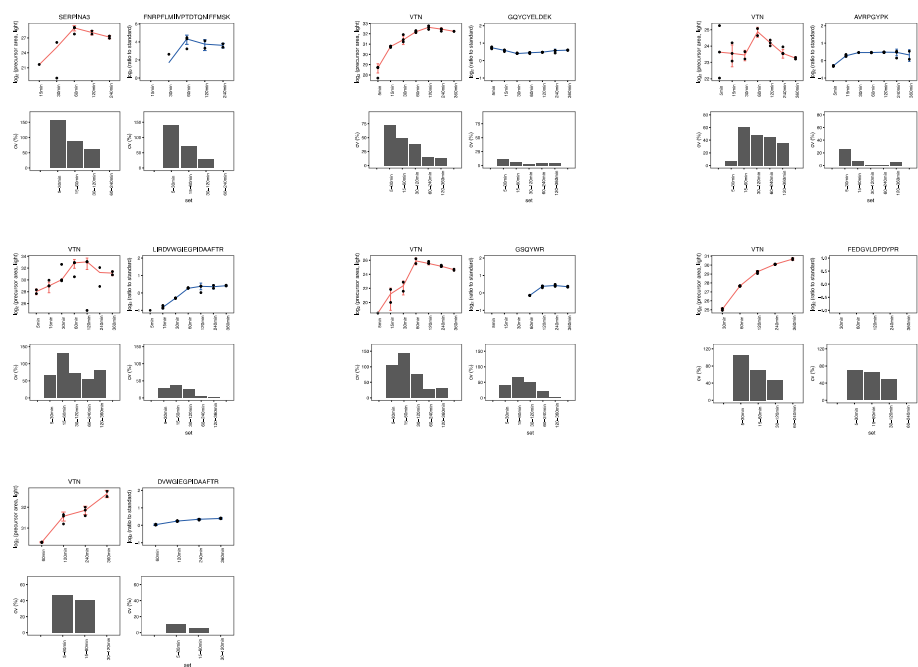




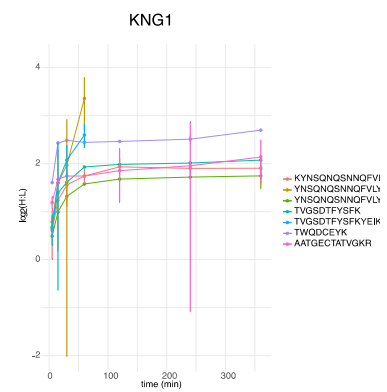
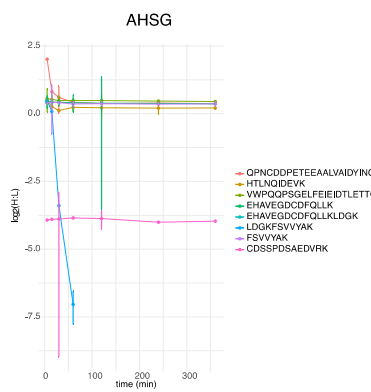
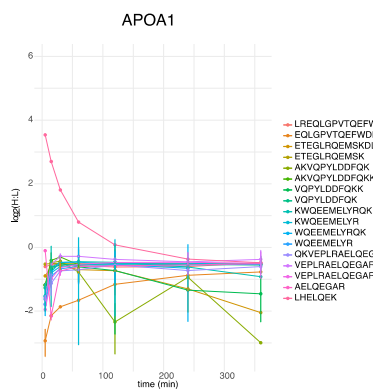
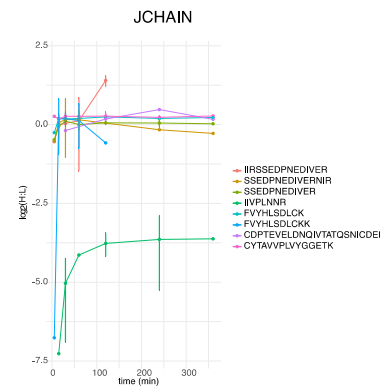
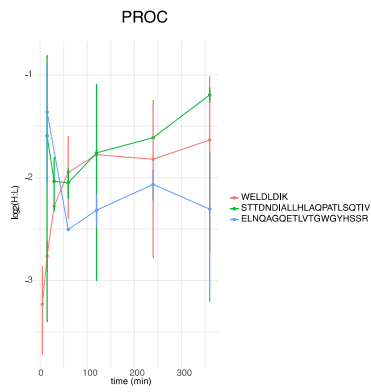
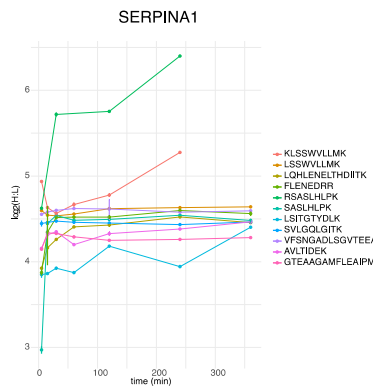
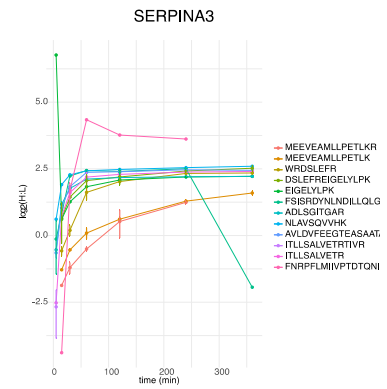
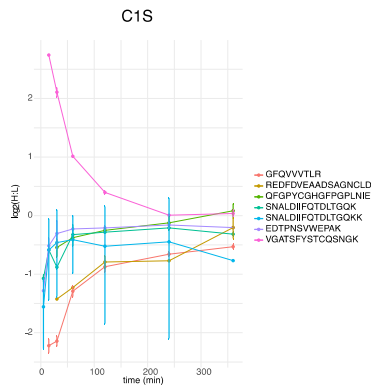
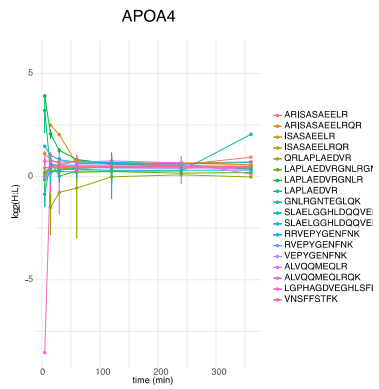


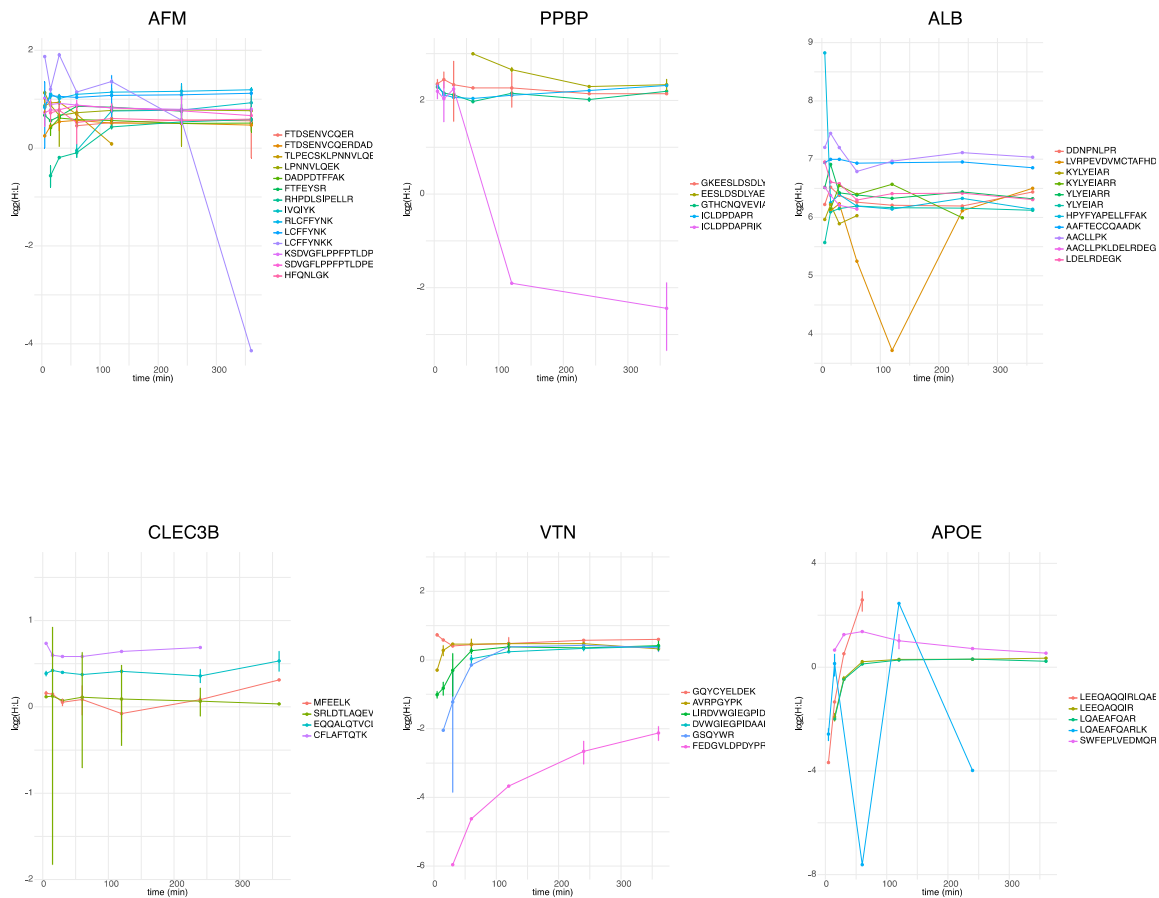






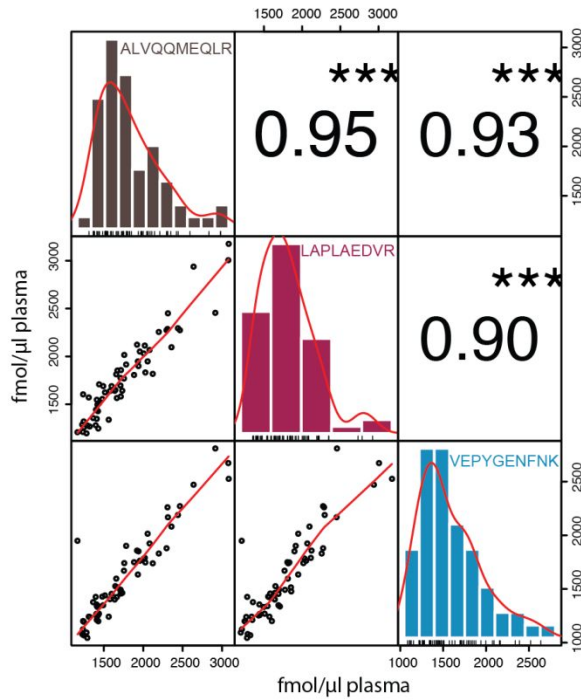
**Figure S-5 – All peptides successfully quantified from the digestion kinetics experiment.** For each peptide; *Top row* - Peptide quantification performed in triplicate, *Red* - Intensity-based quantification. *Blue* - Normalized to standard; *Bottom row* - CV calculated across three time-points for the peptide, *Left*, intensity-based. *Right*, normalized to standard.





**Figure S-6 – Proteins quantified by multiple proteotypic peptides.** Release of tryptic peptides upon digestion is visualized over time (x-axis). Peptide levels have been normalized to spiked in standard (y-axis,  $\log_2$  ratio to standard). Each point is visualized as the median calculated from three technical replicates including error bars (1 sd). Colors represent different proteotypic peptides.





**Figure S-7 – APOA4 peptide correlation across multiple human plasma samples (n = 3).** Three peptides from the same SIL PrEST used to quantify APOA4 across multiple individuals (n = 16) sampled at four different time-points. Histograms across the diagonal show the overall distribution together with scatter plots (pair-wise comparisons) with reported correlation values (Pearson's r). The concentration of the endogenous level is indicated on the y- and x-axis.