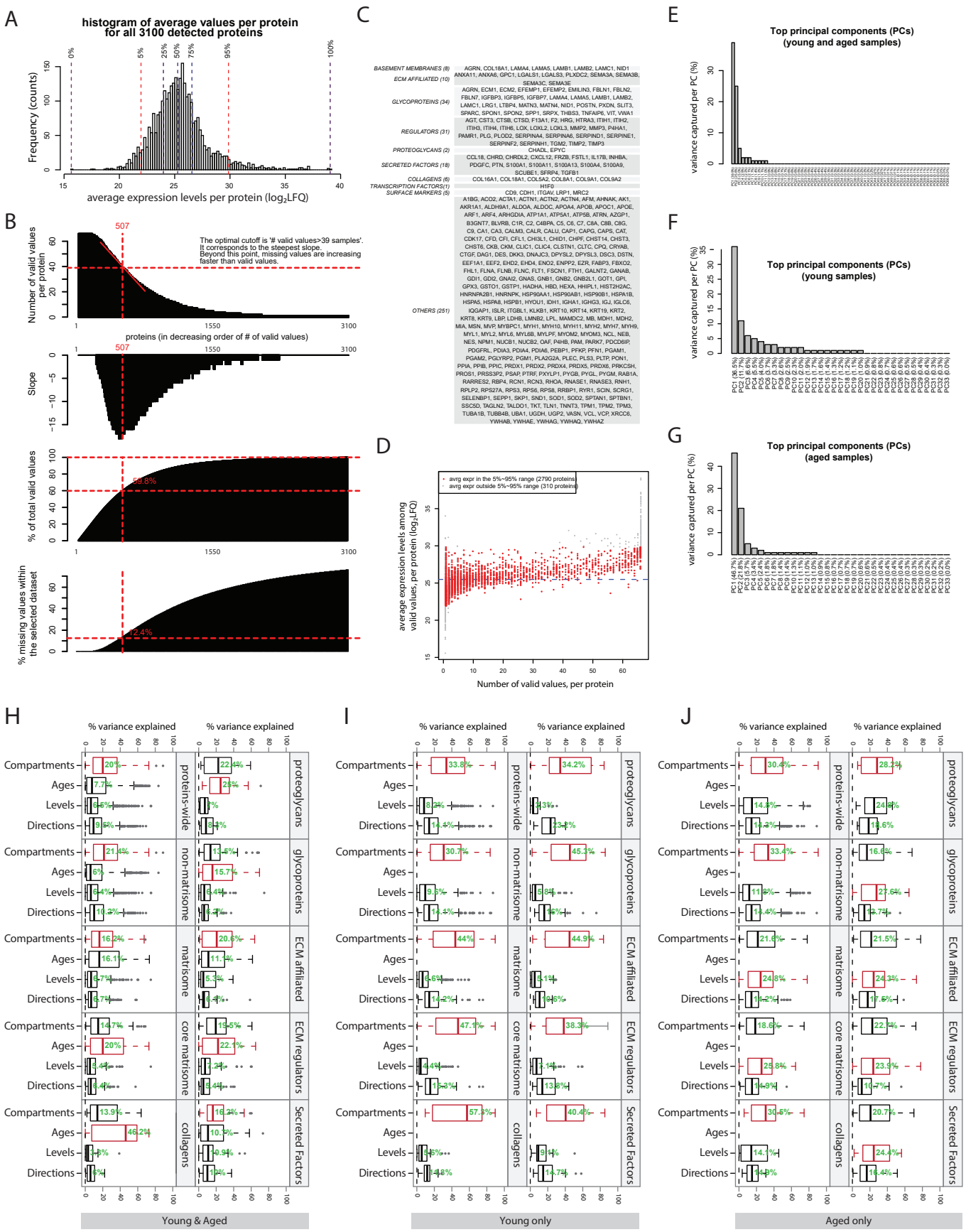

Figure 2—figure supplement 1. Selection of genes for performing PCA and results of ANOVA on assessing global data variability.
(A) Histogram and key percentiles of the average expression levels of all 3100 detected proteins. Average was calculated based on the valid samples of each protein. (B) Selecting an optimal cutoff above which proteins will be used for performing PCA. Upper panel, proteins were ordered in decreasing order by its number of valid values. Second panel: the slope of the upper panel. Third panel: the fraction of all valid values in the whole data-set captured per each cutoff. An optimal cutoff corresponding to the cutoff of '>39 valid values' was found where the slope is steepest. Lower panel: the fraction of missing values within the selected dataset at each cutoff. (C) A list of the proteins (categorized by their functional families) that both meet the criteria in (B) and fall in the 5%~95% range in (A). (D) A scatter plot showing the positive relation between number of valid values per protein, and the average expression level per protein. (E)-(G), the percentages of variance captured by the top principal components (PCs) in the whole dataset (E), young samples only (F) and aged samples only (G). (H)-(J) Horizontal box plots showing the percentages of variance explained by four phenotypic factors, for different scopes of protein sets. (H) young and aged combined. (I) Young only. (J) Aged only.