Constructing a Plasma Nutriproteome for Population Assessment: Analytical Considerations

Hyunju Kim,[1] Kerry Shulze,[1] Casey Rebholz,[1] Gwen Sincerbeaux,[2] Lee Shu-Fane Wu,[1] Sarah Baker,[1] James Yager,[1] Parul Christian,[1] Luigi De Luca,[1] John Groopman,[1] Robert Cole,[3] and Keith West[1]

[1]Johns Hopkins Bloomberg School of Public Health; [2]Johns Hopkins University; and [3]Johns Hopkins School of Medicine

Objectives: Micronutrient status is rarely assessed in low-income settings. Proteomics may offer a proxy by measuring plasma proteins correlated with nutrients on a single platform. However, the proteome is huge, diverse and measured in different ways. We describe an analytic framework and decision process to explore nutrient: protein (N:P) associations for micronutrient status assessment.

Methods: In plasma of 435 1st trimester women in rural Bangladesh, we compared relative protein abundance, revealed by a multiplexed slow off-rate modified aptamer assay (SomaLogic, Inc), to biochemical concentrations of vitamins A, D, E, B_9, B_{12}, Zn, Se, Cu, I, & Fe, 5 carotenoids, cholesterol and AGP. After $\log_2$-transforming protein abundance per convention, N:P relationships were summarized by simple linear regression. We assessed reliability by Pearson correlation (r_p) and coefficients of variation (CV) in 20 blind duplicates. To define each plasma *nutriproteome* [proteins correlating at a false discovery rate < 0.05], in all samples we explored a) normalizing protein abundance to the median of our sample vs not, b) assessing correlations by r_p vs Spearman rank (r_s) estimators, and c) $\log_2$-transforming ($\log_2$) nutrients vs not. We compared differences in the number of proteins and N:P correlative strength (either more negative/positive by r_p or r_s) in each proteome when nutrient concentrations were left untransformed vs $\log_2$-transformed.

Results: In duplicates, $\log_2$-transformed proteins that were normalized, vs not, to the sample-specific median generated higher median r_p (0.92 vs. 0.87) and r_s (0.87 vs. 0.85) and lower CV (4.8% vs. 11.7%). The median (IQR) size of the nutriproteomes was 147 (41–340) proteins by r_p and 87 (29–639) by r_s. For >50% of proteins in 13 nutriproteomes, r_p was stronger than r_s (in either +/− direction), favoring use of r_p. Log_2-transforming folate (B_9), Zn & cholesterol increased proteome size by 39, 223 & 875 proteins and strengthened r_p for >50% of proteins than untransformed nutrients. Other proteomes remained larger when nutrient concentrations remained untransformed.

Conclusions: Comparing plasma protein: nutrient associations via methods of normalization, transformation, and correlation offers a framework to guide plasma nutriproteome definition.

Funding Sources: Johns Hopkins University.