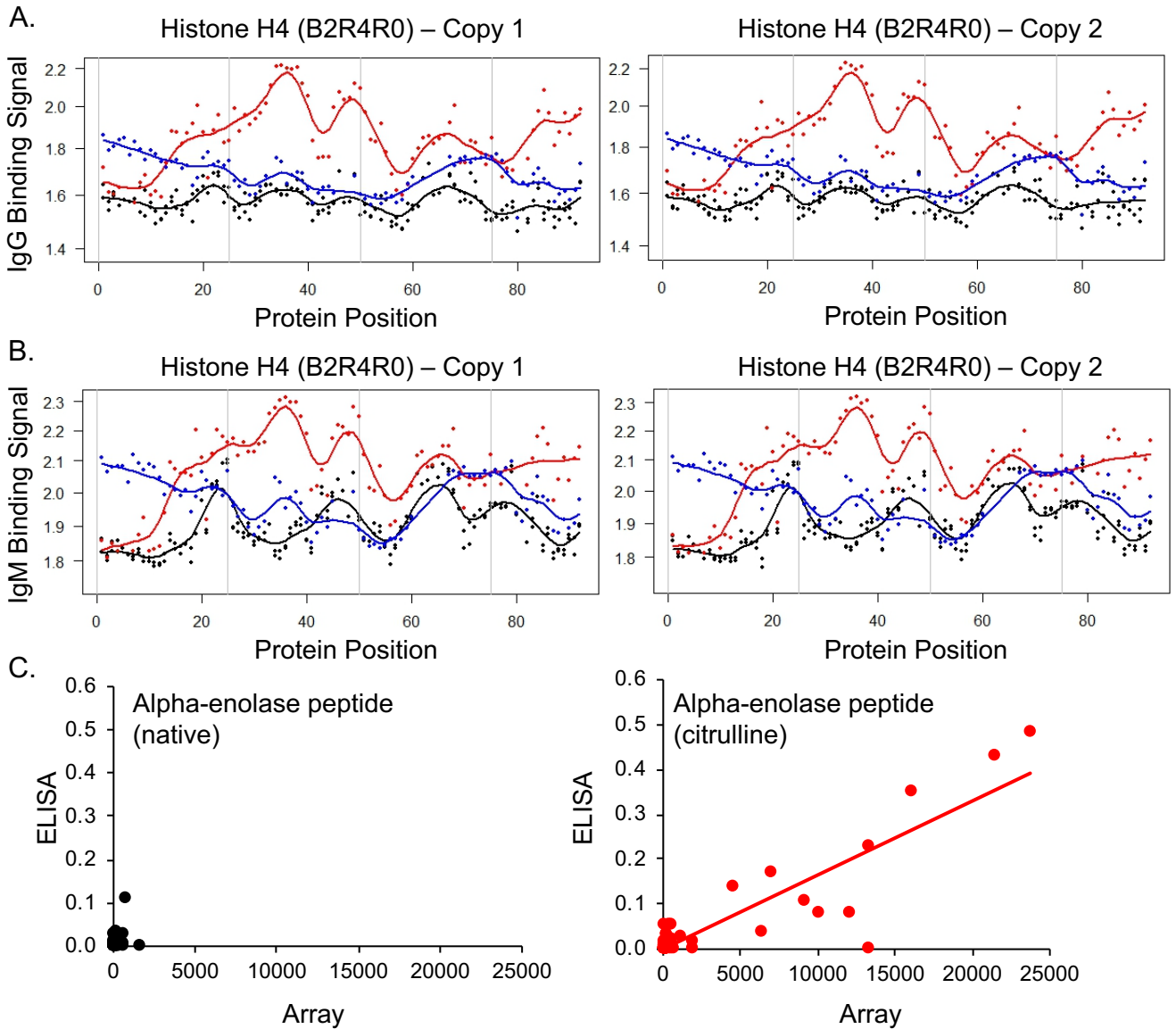
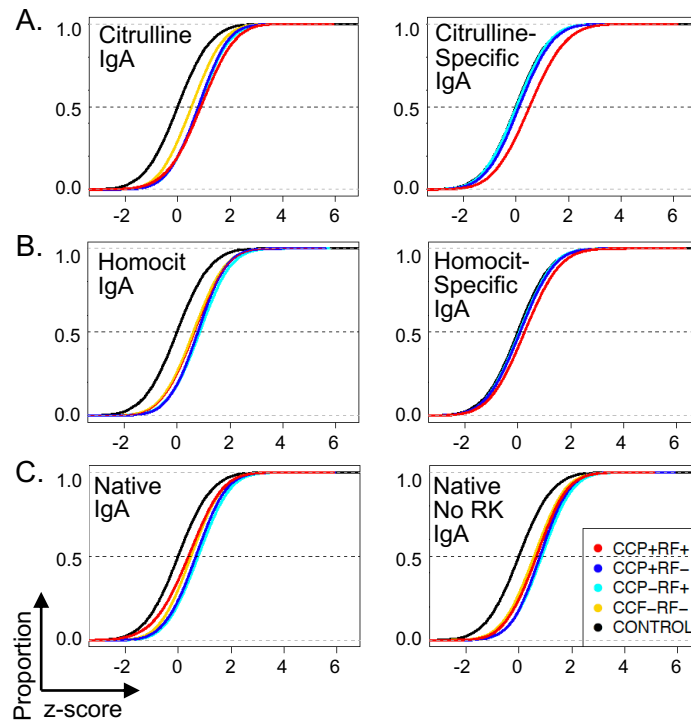


**Supplementary Figure 1. Cluster analysis.** A hierarchical cluster analysis of the full peptide array IgG binding data presents similarities between the high-dimensional abundance profiles using a dendrogram projected into two dimensions as labeled by subject (A) and chip (roman numeral) with numbered array (B). As can be seen in (A), the six serum samples run in duplicate (marked with black dots) on different chips and array positions cluster more closely with each other than with any other subjects, demonstrating a high level of reproducibility between samples and across arrays. Moreover, CCP+RF+ subjects tend to cluster together. In (B), cluster analysis shows no clustering of samples based on array or chip and thus no array-specific bias was seen.



**Supplementary Figure 2. Reproducible results for antibody binding for the high density peptide array.** Raw signal intensity on a double log scale for IgG (A) and IgM (B) binding is plotted for each native (black), citrulline-containing (red), and homocitrulline-containing (blue) peptide according to its position in histone H4. Left panel is one copy of histone H4 peptides and right panel is an exact duplicate of histone H4 peptides, with similar binding for each copy. Similar results were seen for histone H1, H2A, H2B, and H3 (data not shown). C. IgG binding signal by ELISA for each subject for a native (left) or citrulline-containing (right) peptide of alpha-enolase starting at amino acid 10 is plotted against the IgG binding signal as determined by the array. For citrulline-containing peptides: Pearson correlation 0.89,  $p < 0.0001$ .



**Supplementary Figure 3. Broad patterns of IgA reactivity to peptides in rheumatoid arthritis.** IgA binding was measured for all samples for each peptide and the z score for each peptide calculated for each rheumatoid arthritis group (CCP+RF+, CCP+RF-, CCP-RF+, and CCP-RF-) compared to control. The proportion (y axis) of z scores less than the specific z value on the x axis was graphed as an empirical cumulative distribution function (ecdf) plot. The median z score for each rheumatoid arthritis group is where its ecdf curve crosses the dotted line at 0.5. A. Citrulline-containing peptides and citrulline-specific binding (binding values for citrulline-containing peptides minus corresponding arginine-containing peptides), n=35,459 peptides or peptide pairs. B. Homocitrulline-containing peptides and homocitrulline-specific binding (binding values for homocitrulline-containing minus corresponding lysine-containing peptides), n=41,608 peptides or peptide pairs. C. All native peptides and native peptides excluding peptides with arginine or lysine, n=95,761 native peptides, 18,694 native peptides without arginine or lysine. For all panels, n=12 samples per group.