



Fig 5. Context-based measurements versus the network distance in the E. coli metabolic network. Figure shows (a). Gene phylogenetic Pearson correlation. (b). Gene co-expression Pearson correlation. (c). Gene clustering score and (d) Protein fusion events. In general, context-based scores decrease monotonically as the metabolic distance between two genes increases. Gene clustering scores were calculated as  $\log_{10}(\text{clustering probability})$ . The larger the score, the more likely two genes cluster. Protein fusion events were calculated as the number of pairs of genes fused divided by total number of gene pairs at a certain metabolic distance.