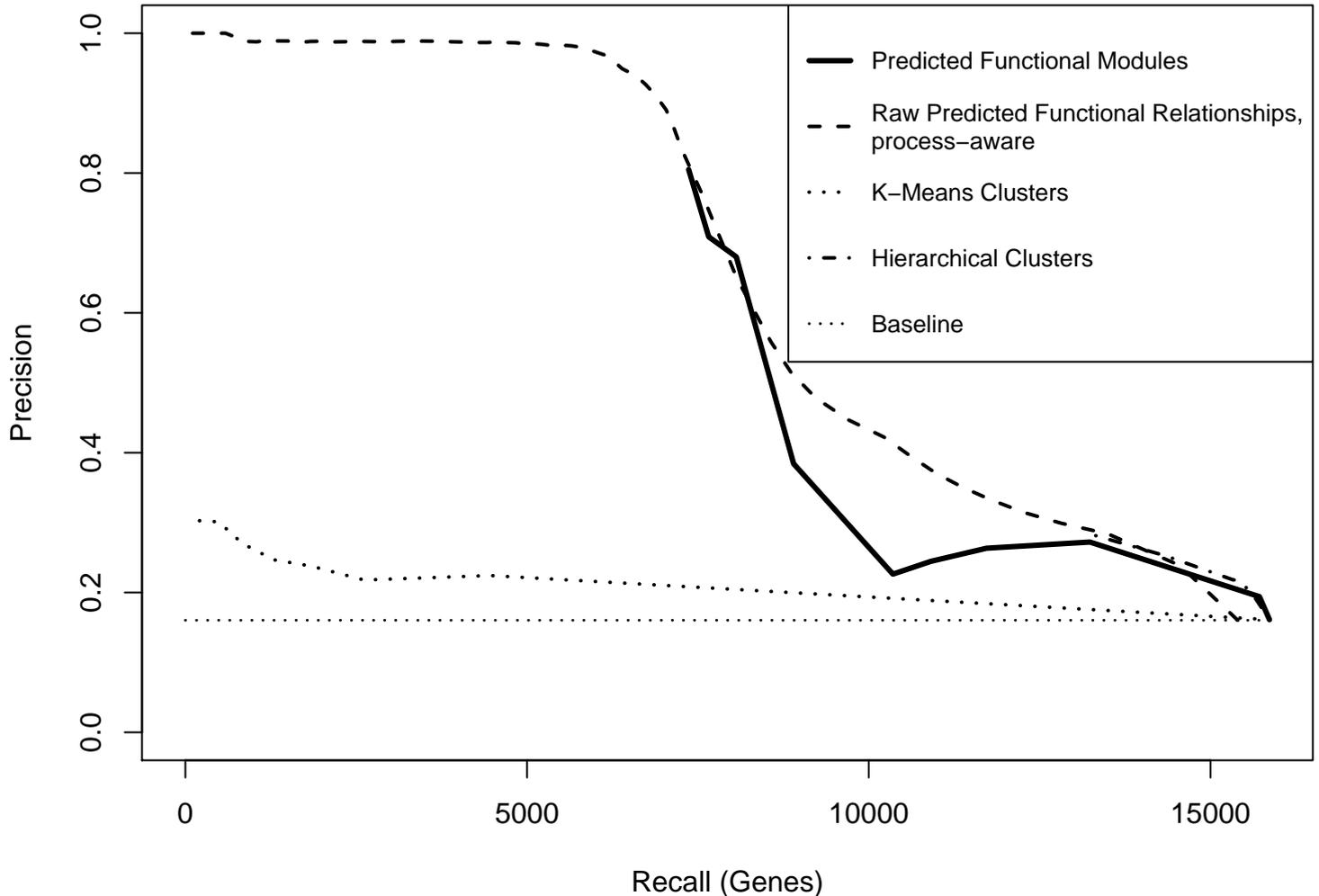# Performance in recapitulating known biology

Supplemental Figure 3: Performance of predicted functional modules in recapitulating known biology. To validate the functional specificity of the hierarchical modules predicted by our data integration process, we analyzed their ability to recover information from a held−out portion of our gold standard (as per Figure 1B's evaluation of individual predicted functional relationships). These modules, a hierarchy of gene clusters drawn from our functional relationship networks (not from usual clustering metrics such as distance or correlation), were compared with other clustering schemes (k−means with k=5, 10, 50, 100, 500, 1000, and 5000, and hierarchical clustering cut to produce 5, 10, etc. clusters). In all cases, more specific parameters (modules lower in the hierarchy, larger k, or smaller diameters) are expected to produce more precise clusters in which fewer genes cocluster. This is reflected in the plot above, with our predicted functional modules demonstrating a substantial advantage over less structured clusterings (all of which consume our predicted functional relationship networks as input). The hierarchical structure and discrete sets of related genes that make up the functional modules can be an advantage over the raw, continuous predicted functional relationship scores for some applications; these results show that the functional modules do not sacrifice substantial biological accuracy in order to offer these features.