

**Figure S4:** Average silhouette coefficient vs. HAC model. Silhouette coefficient (SC) over a range of possible number of clusters ( $k$ ). SC is maximized for  $k = 3$  and  $k = 4$ . Since  $k = 3$  is concordant with the metadata, this HAC model was preferred.

