



Fig. 1. Determining the dimensionality of a clustering problem. (A and B) Representation of the mRNA clustering problem consisting of >14,000 mRNAs measured across 89 cell lines. Data are from Lu *et al.* (6). When the mRNAs are clustered, the mRNAs are the objects and each cell line represents a feature, resulting in an 89-dimensional problem (A). When attempting to classify normal and tumor cell lines using gene expression, the cells lines are the objects and each mRNA is a feature, resulting in a clustering problem with thousands of dimensions (B). (C) Effect of dimensionality on sparsity. (D) Effect of dimensionality on coverage of the data based on SD from the mean.