## First two components colored by cohort First two components colored by ER status Only 1000 training samples Only 1000 training samples 5 PC2, 13.24% variation PC2, 13.24% variation cohort er\_status neg pos metabric tcga -5 -5 5 5 PC1, 19.21% variation PC1, 19.21% variation Second and third components colored by cohort Second and third components colored by ER status Only 1000 training samples Only 1000 training samples PC3, 5.55% variation PC3, 5.55% variation cohort er\_status metabric neg pos tcga -5 5 PC2, 13.24% variation PC2, 13.24% variation Second and third components colored by PAM50 Only 1000 training samples PC3, 5.55% variation basal claudin-low her2 luma lumb nc normal

-5

PC2, 13.24% variation