

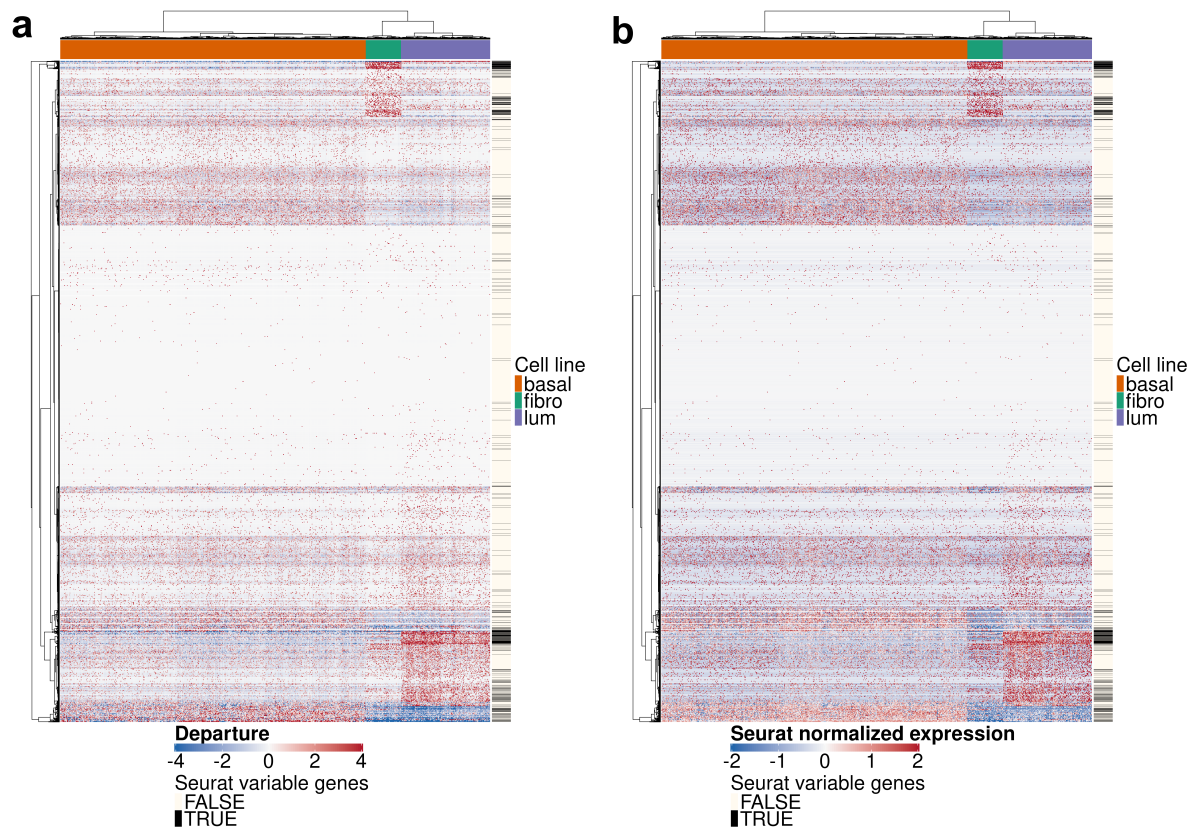


Figure S2: A heatmap view of the data representations of a mixture cell lines data set (three cell lines mixture data). Data representations based on (a) DIPD and (b) Seurat normalized and scaled counts before feature selection. The black colored lines in the sidebars on the right represent the top 2,000 most variable genes kept by the Seurat pipeline. Visually, both data representations effectively demonstrate the differentially expressed genes among the three cell lines. However, highly expressed genes within single cells, as depicted by the bright red spots, may potentially play a role in clustering but many are filtered out by Seurat.