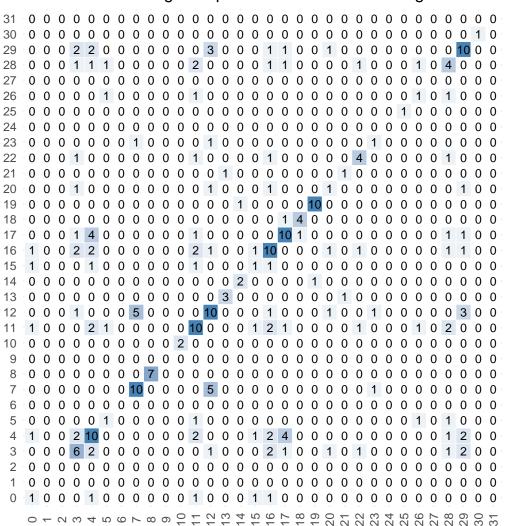Number of similar genes per cluster and considering confounders