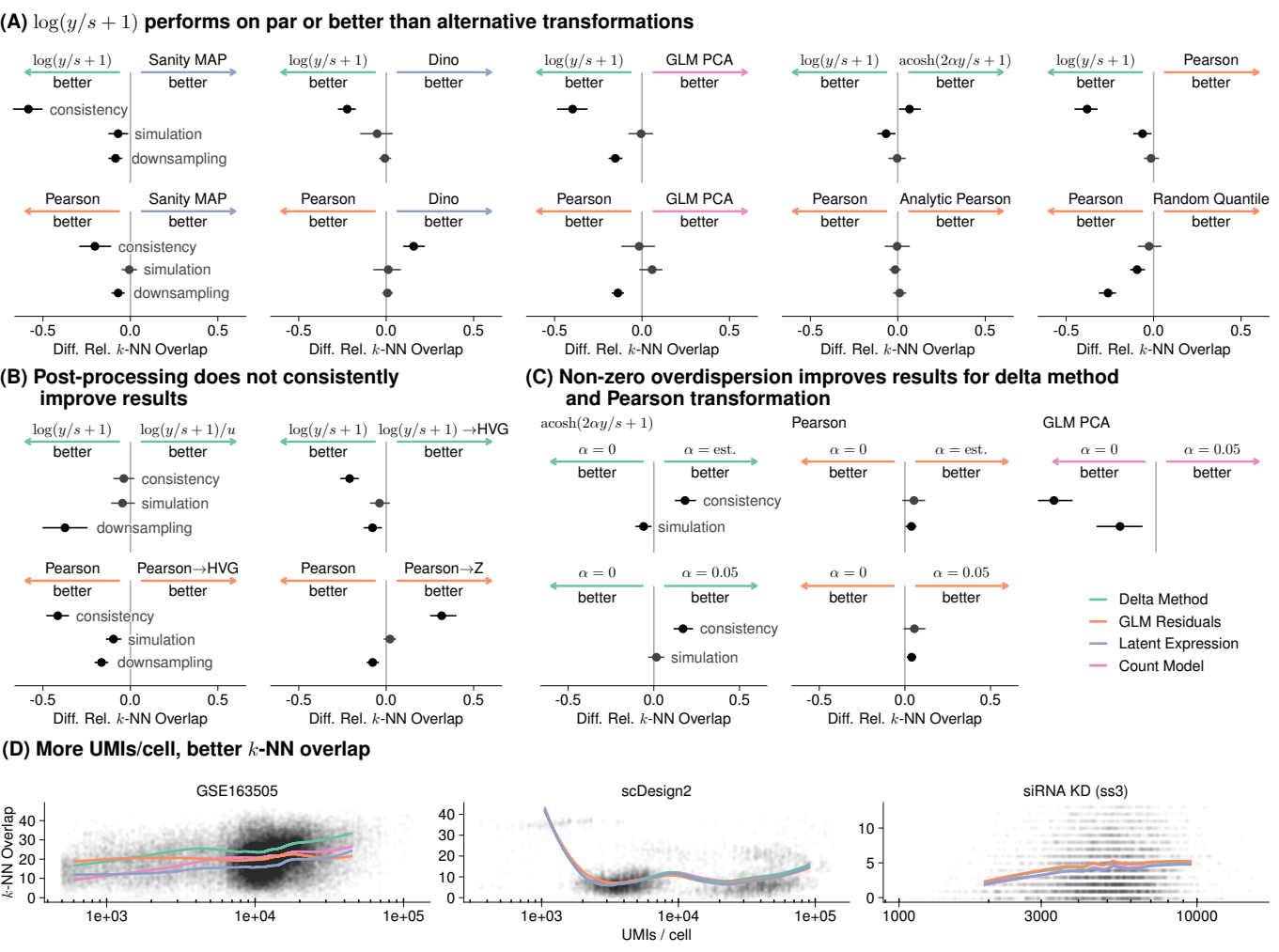
(A) $\log(y/s+1)$ performs on par or better than alternative transformations
$\log(y/s+1)$ better
Sanity MAP better
Dino better
GLM PCA better
$\text{acosh}(2\alpha y/s+1)$ better
Pearson better
Analytic Pearson better
Random Quantile better
consistency
simulation
downsampling
-0.5
0.0
0.5
Diff. Rel. k-NN Overlap
(B) Post-processing does not consistently improve results
$\log(y/s+1)/u$ better
$\log(y/s+1)\to$HVG better
Pearson→HVG better
Pearson→Z better
(C) Non-zero overdispersion improves results for delta method and Pearson transformation
$\alpha=0$ better
$\alpha=$ est. better
$\alpha=0.05$ better
Delta Method
GLM Residuals
Latent Expression
Count Model
(D) More UMIs/cell, better k-NN overlap
GSE163505
scDesign2
siRNA KD (ss3)
k-NN Overlap
UMIs / cell
1e+03
1e+04
1e+05
1000
3000
10000