(A) Consistency: 10X gene subset 1 vs. gene subset 2 GSE130931 (#PCs=50) GSE142647 (#PCs=50) GSE163505 (#PCs=50) GSE150068 (#PCs=50) GSE158941 (#PCs=50) $\log(y/s + 1) -$ $\operatorname{acosh}(2\alpha y/s + 1) -$ $\log(y/s + 1/(4\alpha)) -$ $\log(CPM + 1) -$ Not the second seDelta Method $\begin{array}{c} \log(x) & \text{H} + 1) \\ \log(y/s+1) & \text{HVG} \\ \log(y/s+1) \rightarrow \text{HVG} \\ \log(y/s+1) \rightarrow \text{Z} \\ \log(y/s+1) \rightarrow \text{HVG} \rightarrow \text{Z} \end{array}$ Pearson sctransform Analytic Pearson Random Quantile Pearson (no clip) Pearson → HVG Pearson → Z Pearson → Z Residuals GLM Sanity MAP Sanity Distance Dino Normalisr Exp. Ę. **GLM PCA** GSE164017 (#PCs=50) GSE178765 (#PCs=50) GSE179714 (#PCs: =50) GSE179831 (#PCs=50) GSE184806 (#PCs=50) $\begin{array}{c} \log(y/s+1) \\ \operatorname{acosh}(2\alpha y/s+1) \\ \log(y/s+1/(4\alpha)) \\ \log(\operatorname{CPM}+1) \\ \log(y/s+1) \\ \log(y/s+1) \to \operatorname{HVG} \\ \log(y/s+1) \to \operatorname{TVG} \to \operatorname{CM}(y/s+1) \\ \end{array}$ Pearson sctransform Analytic Pearson Random Quantile **GLM Residuals** Pearson (no clip) Pearson→HVG Pearson→Z Pearson→HVG→Z Sanity MAP Sanity Distance Dino Normalisr Ľ GLM PCA NewWave 10 20 300 k-NN Overlap (B) Simulation: Ground truth vs. simulated counts Dyngen (#PCs=5) Linear Walk (#PCs=10) muscat (#PCs=10) Random Walk (#PCs=200) scDesign2 (#PCs=50) Delta Method Pearson sctransform Analytic Pearson Random Quantile Pearson (no clip) Pearson→HVG Pearson→Z Pearson→HVG→Z **GLM Residuals** Sanity MAP Sanity Distance Dino Normalisr GLM PCA NewWave k-NN Overlap (C) Downsampling: Original vs. downsampled Smart-seq3 data HEK (ss3) (#PCs=10) mcSCRB (#PCs=10) siRNA KD (ss3) (#PCs=50) Fibroblasts (ss3) (#PCs=10) Fibroblasts 2 (ss3) (#PCs=10)

