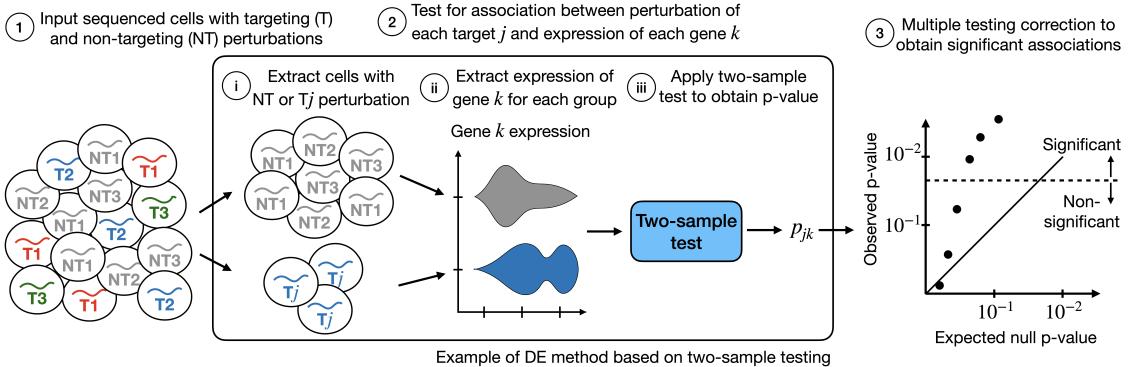


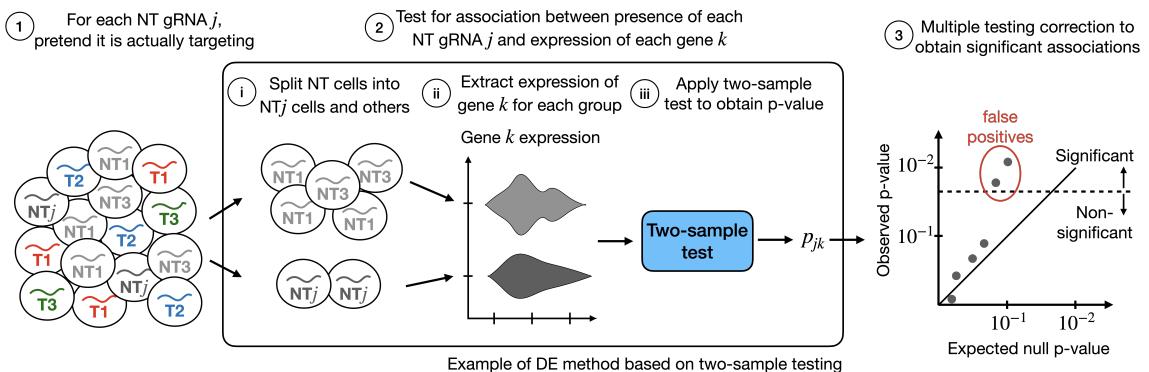
Robust differential expression testing for single-cell CRISPR screens

Timothy Barry¹, Kaishu Mason³, Kathryn Roeder^{1,2}, and Eugene Katsevich³

a Typical differential expression analysis for single cell CRISPR screens

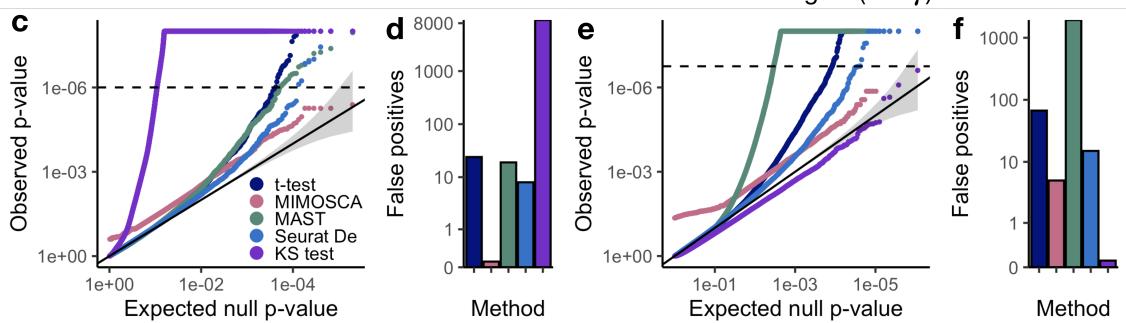


b Undercover differential expression analysis for calibration assessment



Papalex data

Frangieh (IFN- γ) data



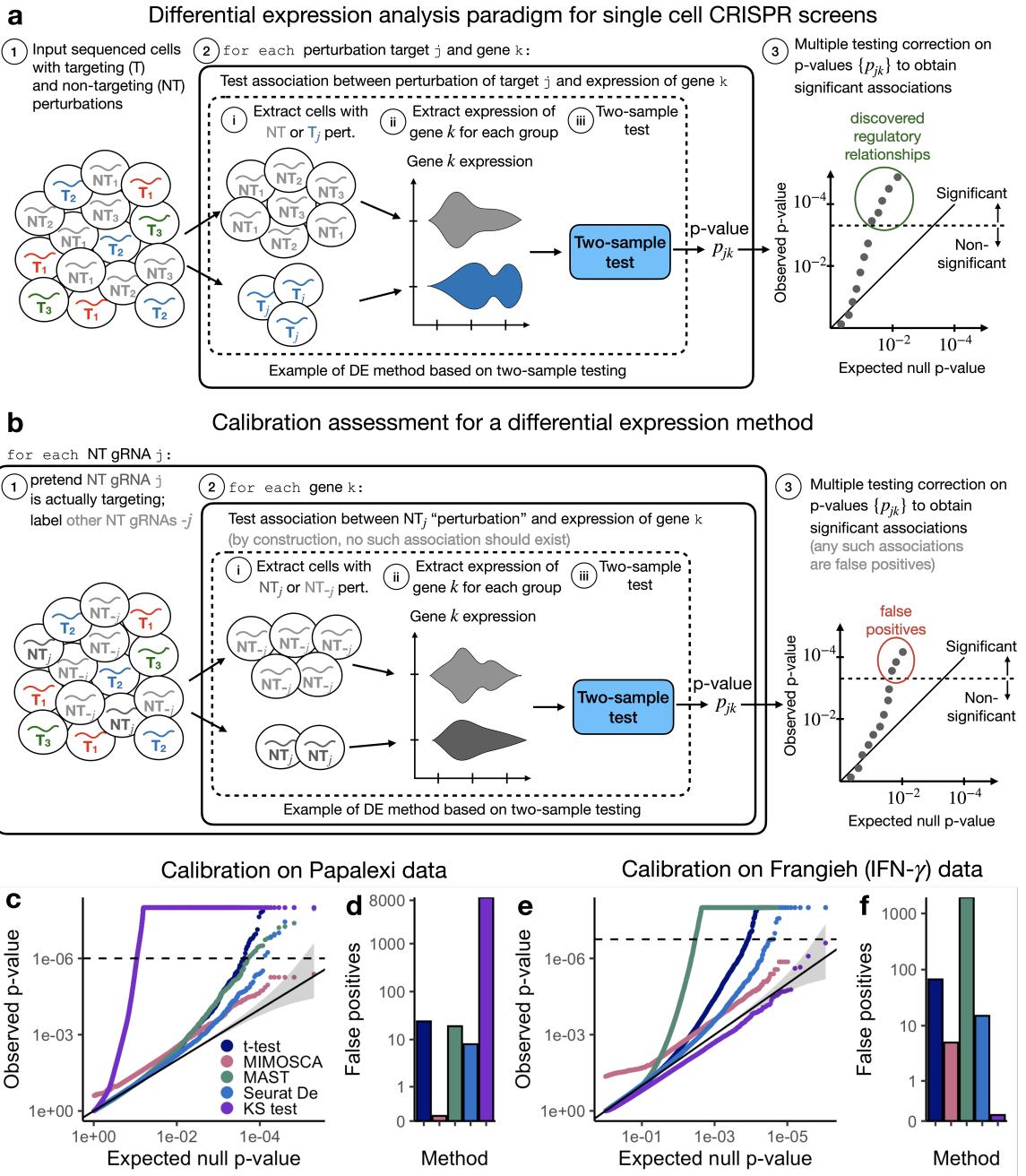


Figure 1: Insert caption here.

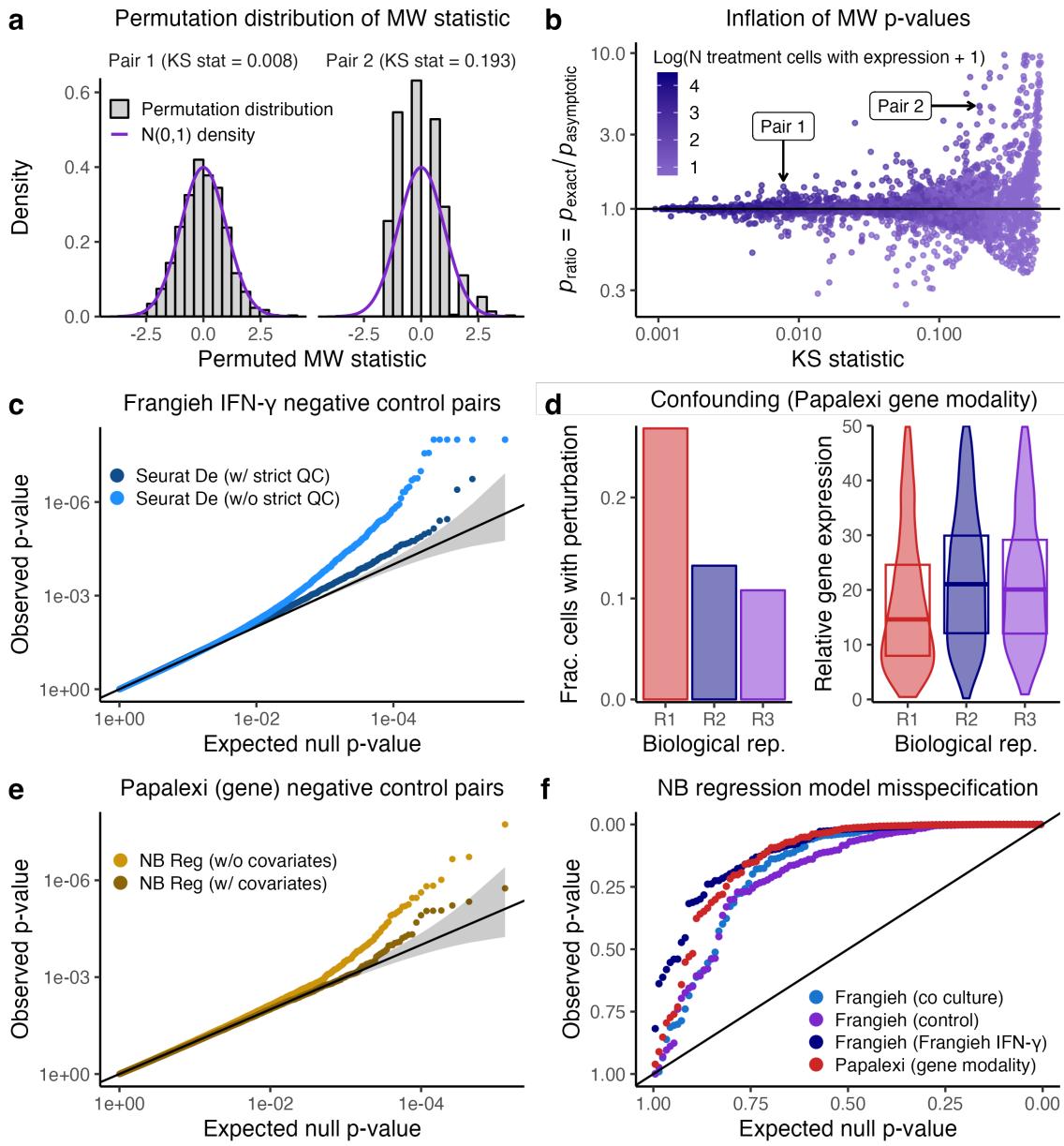


Figure 2: Sparsity, confounding, and model misspecification are core analysis challenges in single-cell CRISPR screen analysis. **a**, The exact null distribution of the Mann-Whitney (MW) test statistic (obtained via permutations; grey) on two pairs from the Frangieh IFN- γ data. The MW test (and thus Seurat DE) approximates the exact null distribution using a standard Gaussian density (purple). For pair 1 (left), the Gaussian

approximation to the exact null distribution is good (KS statistic = 0.008); for pair 2, by contrast (right), the approximation is poor (KS statistic = 0.193). **b**, A plot of p_{ratio} (defined as the ratio of the exact MW p-value, p_{exact} , to the asymptotic MW p-value, $p_{\text{asymptotic}}$) vs. goodness of fit of the Gaussian distribution to the exact null distribution (as quantified by the KS statistic). Each point represents a gene-gRNA pair; pairs 1 and 2 (from panel **a**) are annotated. As the KS statistic increases (indicating worse fit of the Gaussian distribution to the exact MW null distribution), p_{ratio} deviates more from one, indicating miscalibration. Points are colored according to the effective sample size (as quantified by the number of treatment cells with nonzero expression) of the corresponding pair. **c**, An application Seurat DE to the IFN- γ negative control data with and without stringent QC; applying stringent QC in this context amounts to filtering for pairs with a very large effective sample size. **d**, An example of confounding on the Papalexie data. Left (resp. right), the fraction of cells that received a given NT gRNA (resp., the relative expression of a given gene) across biological replicates “R1,” “R2,” and “R3.” If we failed to account for biological replicate, we would conclude (incorrectly) the the NT gRNA *decreases* the relative expression of the gene. **e**, Application of NB regression to the Papalexie data. Inclusion of confounders (such as biological replicate) in the regression model improves calibration (although further improvements are possible). **f**, A QQ-plot of p-values obtained from testing for goodness of fit of the NB regression model to the gene expression data (points colored by dataset). The p-values are inflated, indicating that the NB regression model provides a poor fit to some subset of the genes.

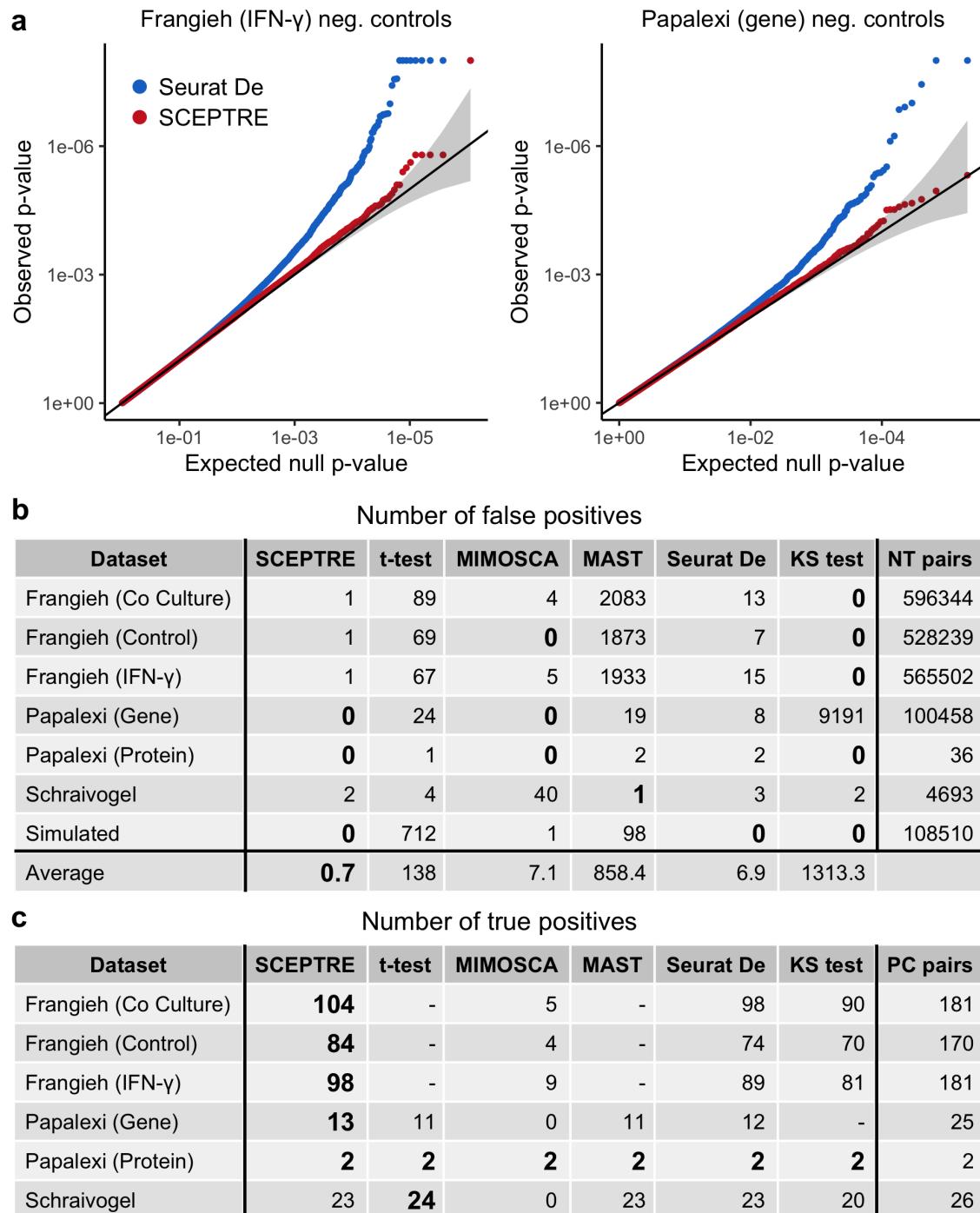


Figure 3: Insert caption here.