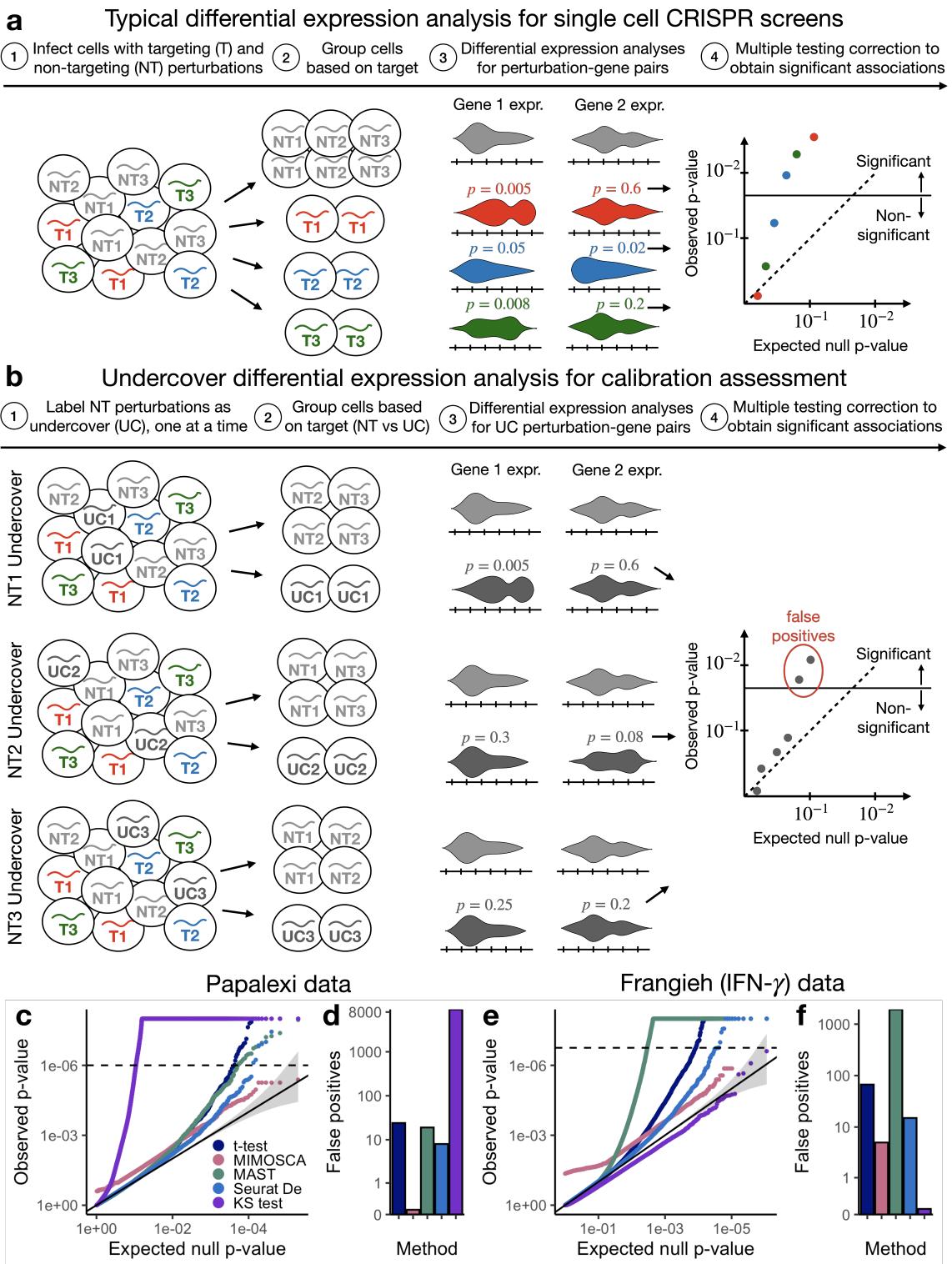
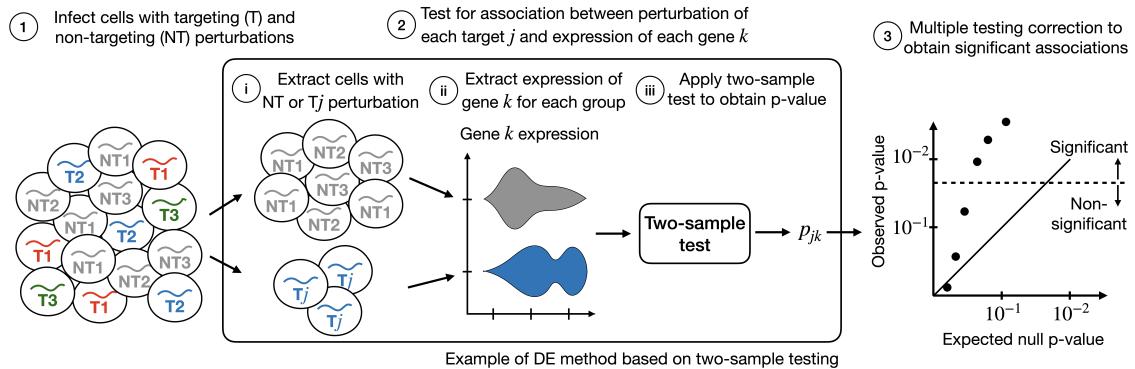


# **Robust differential expression testing for single-cell CRISPR screens**

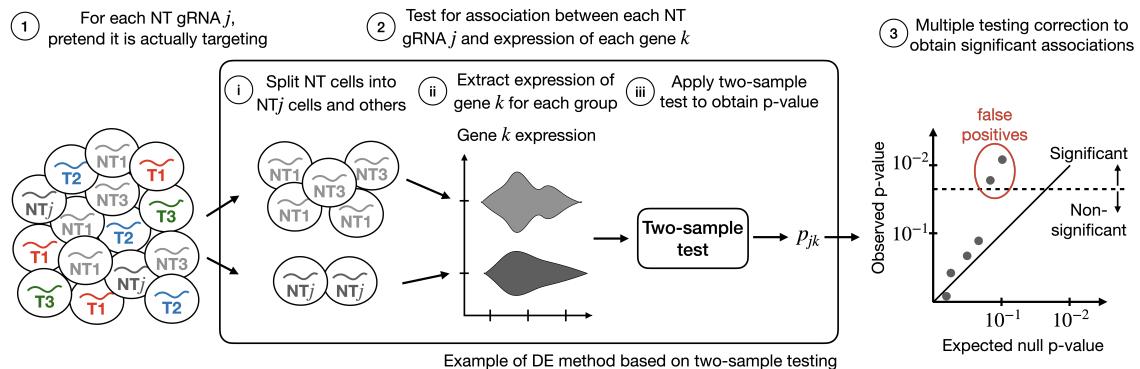
Timothy Barry<sup>1</sup>, Kathryn Roeder<sup>1,2</sup>, and Eugene Katsevich<sup>3</sup>



### a Typical differential expression analysis for single cell CRISPR screens



### b Undercover differential expression analysis for calibration assessment



Papalexi data

Frangieh (IFN- $\gamma$ ) data

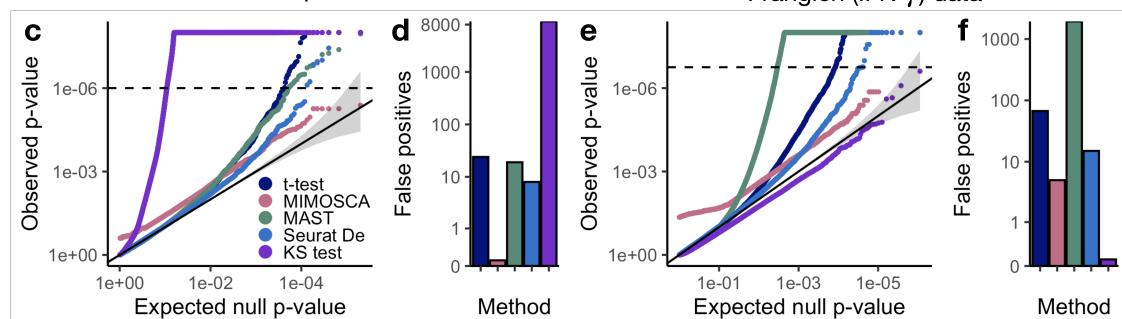
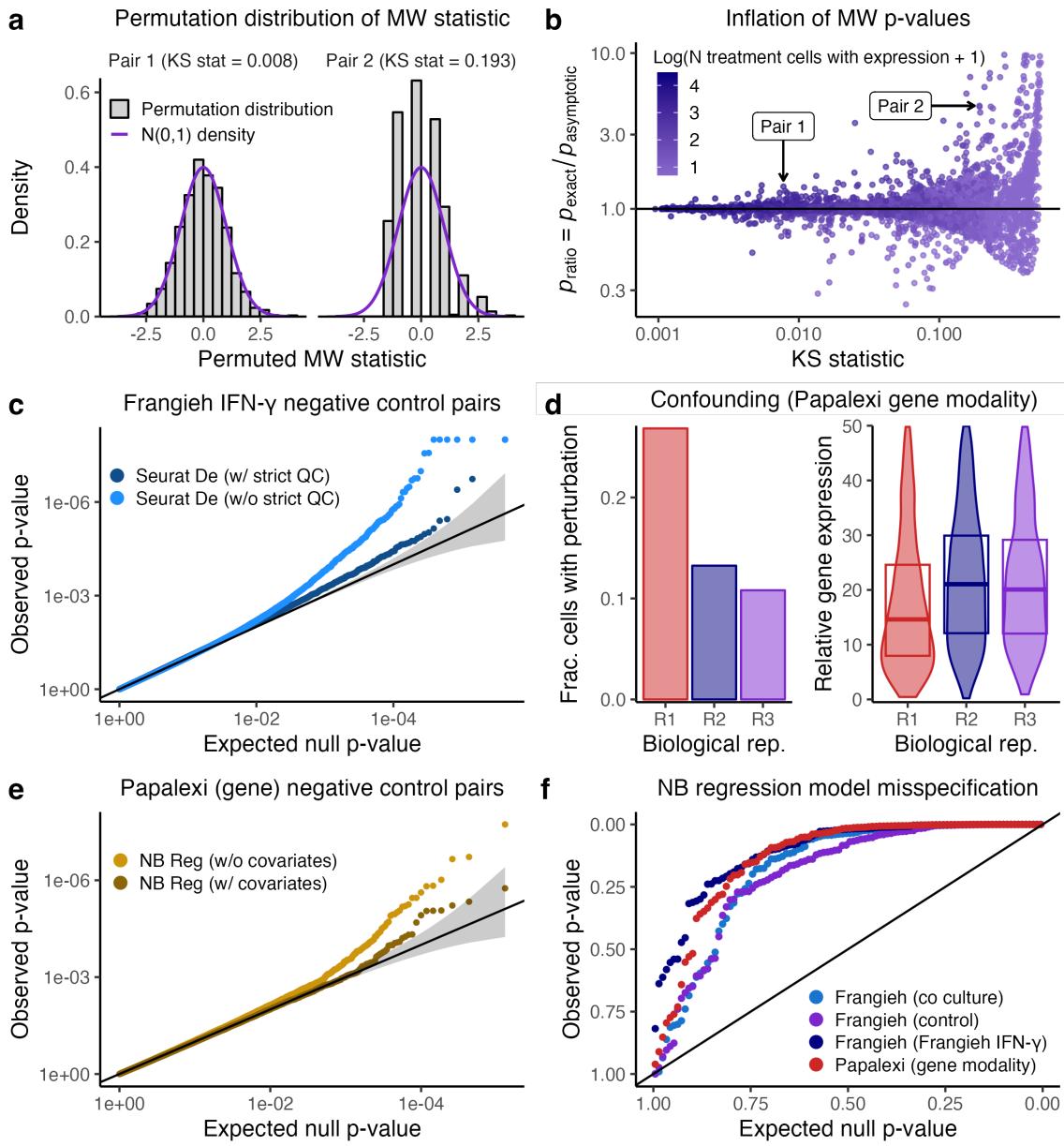


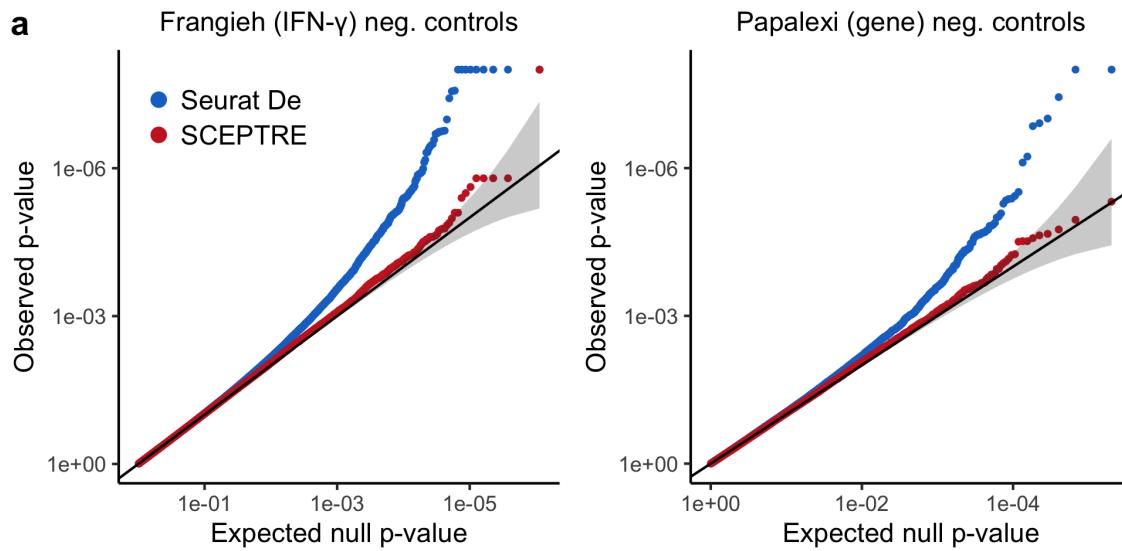
Figure 1: Insert caption here.



**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- $\gamma$  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  $p_{\text{ratio}}$  (defined as the ratio of the exact MW p-value,  $p_{\text{exact}}$ , 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),  $p_{\text{ratio}}$  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- $\gamma$  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.

---



**b** Number of false positives

Dataset	SCEPTR	t-test	MIMOSCA	MAST	Seurat De	KS test	NT pairs
Frangieh (Co Culture)	1	89	4	2083	13	<b>0</b>	596344
Frangieh (Control)	1	69	<b>0</b>	1873	7	<b>0</b>	528239
Frangieh (IFN-γ)	1	67	5	1933	15	<b>0</b>	565502
Papalex (Gene)	<b>0</b>	24	<b>0</b>	19	8	9191	100458
Papalex (Protein)	<b>0</b>	1	<b>0</b>	2	2	<b>0</b>	36
Schraivogel	2	4	40	<b>1</b>	3	2	4693
Simulated	<b>0</b>	712	1	98	<b>0</b>	<b>0</b>	108510
Average	<b>0.7</b>	138	7.1	858.4	6.9	1313.3	

**c** Number of true positives

Dataset	SCEPTR	t-test	MIMOSCA	MAST	Seurat De	KS test	PC pairs
Frangieh (Co Culture)	<b>104</b>	-	5	-	98	90	181
Frangieh (Control)	<b>84</b>	-	4	-	74	70	170
Frangieh (IFN-γ)	<b>98</b>	-	9	-	89	81	181
Papalex (Gene)	<b>13</b>	11	0	11	12	-	25
Papalex (Protein)	<b>2</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>2</b>	2
Schraivogel	23	<b>24</b>	0	23	23	20	26

Figure 3: Insert caption here.