sceptre project update:
August 2022

Tim Barry
Gene Katsevich

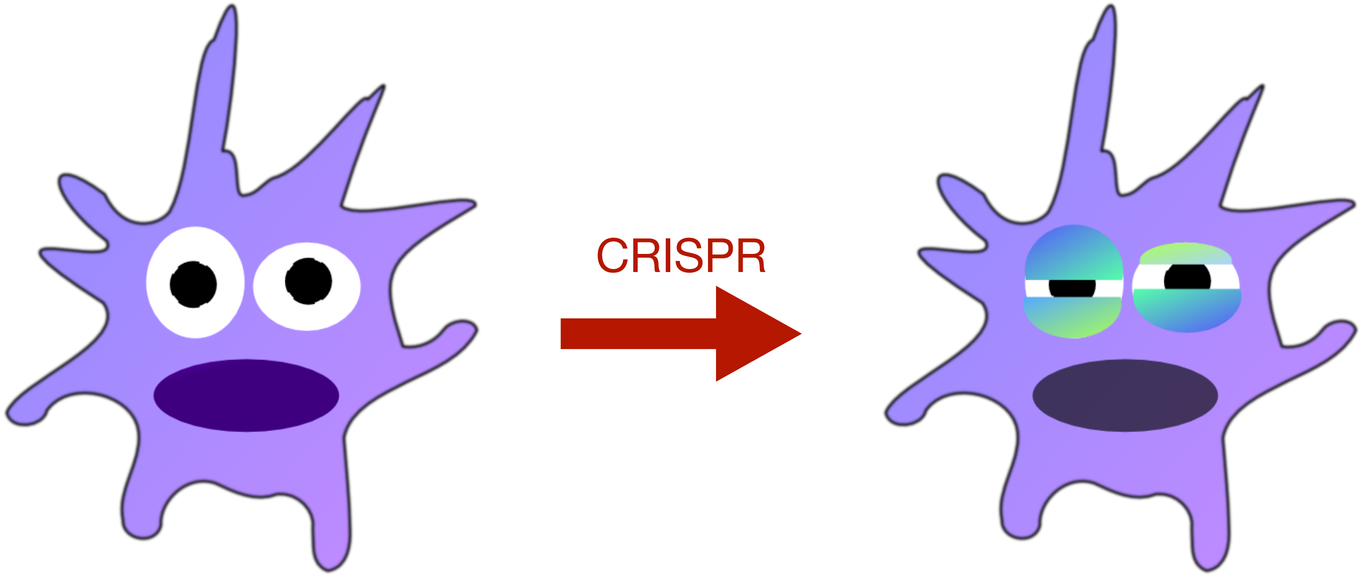
CRISPR is a genome engineering technology.

- Fix genes that cause diseases in humans.
- Transform elephants into woolly mammoths!?
- Accelerate biological discovery.

CRISPR

CRISPR accelerates biological discovery.

1. Identify a region of the genome (e.g., a gene) with unknown function.
2. Perturb this region of the genome with CRISPR.
3. See what happens!



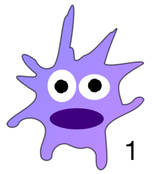
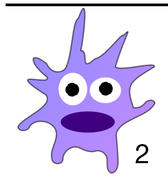
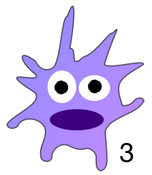
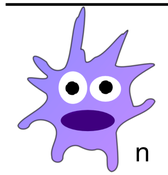
Normal cell

CRISPR-edited cell

What might be the function of the gene *targeted* by CRISPR?

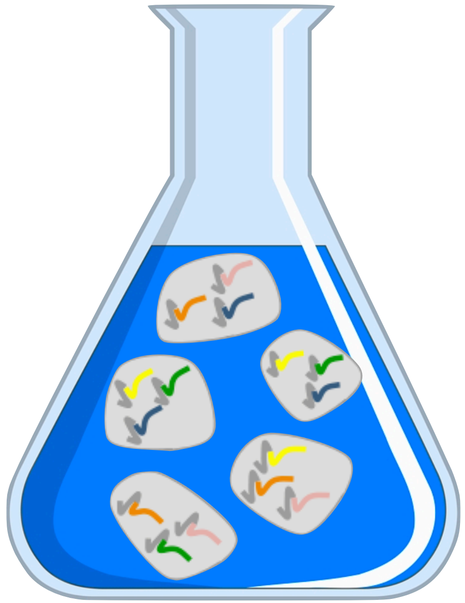
Single-cell RNA sequencing is a technology for measuring gene expressions in individual cells.

i.i.d. cells

	Gene 1	Gene 2	Gene 3		Gene p
	1	0	4	...	2
	0	1	0	...	0
	3	0	0	...	2
	⋮	⋮	⋮		⋮
	0	2	4	...	1

Single-cell CRISPR screens couple CRISPR to single-cell sequencing, enabling scientists to interrogate the effects of perturbations in individual cells.

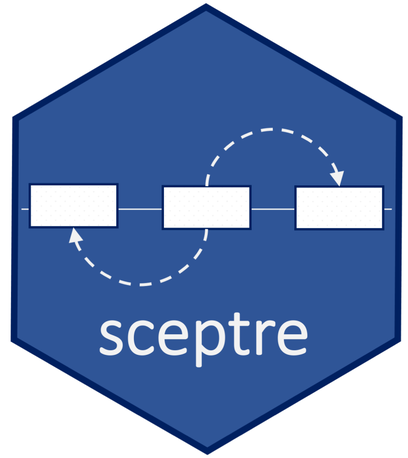
Insert CRISPR perturbations into cells



Single-cell sequencing

- Single cell CRIPSR screens likely will transform biology and medicine.
- However, these screens pose major statistical and computational challenges.

sceptre aims to power statistically sound and computationally efficient single-cell CRISPR screen analysis.



Statistical goals

✅ Control false positives

✅ Make lots of discoveries

Computational goals

✅ Fast and lightweight implementation

✅ Scale to large data

Roadmap

1. sceptre in high MOI

2. sceptre in low MOI

METHOD Open Access

SCEPTRE improves calibration and sensitivity in single-cell CRISPR screen analysis

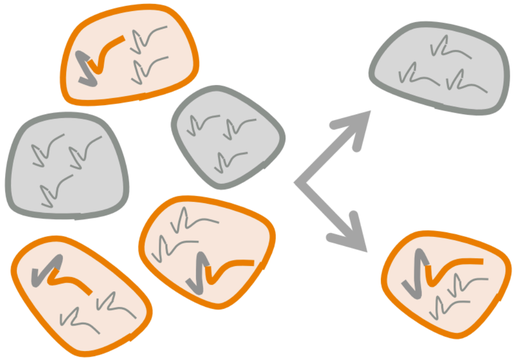
Timothy Barry[1], Xuran Wang[1], John A. Morris[2,3], Kathryn Roeder[1,4] and Eugene Katsevich[5*]

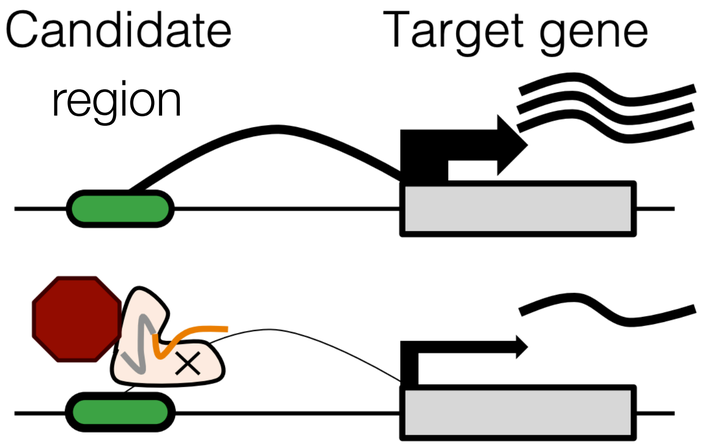
*Correspondence:
ekatsevi@wharton.upenn.edu
[5]Department of Statistics and Data Science, Wharton School, University of Pennsylvania, Philadelphia, PA 19104, USA
Full list of author information is available at the end of the article

Abstract

Single-cell CRISPR screens are a promising biotechnology for mapping regulatory elements to target genes at genome-wide scale. However, technical factors like sequencing depth impact not only expression measurement but also perturbation detection, creating a confounding effect. We demonstrate on two single-cell CRISPR screens how these challenges cause calibration issues. We propose SCEPTRE: analysis of single-cell perturbation screens via conditional resampling, which infers associations between perturbations and expression by resampling the former according to a working model for perturbation detection probability in each cell. SCEPTRE demonstrates very good calibration and sensitivity on CRISPR screen data, yielding hundreds of new regulatory relationships supported by orthogonal biological evidence.

High MOI single-cell CRISPR screens involve inserting **multiple** gRNAs into each cell.





1. For a given perturbation (orange), partition the cells into two groups: perturbed and unperturbed.

2. For a given gene, perform a differential expression analysis across these two groups of cells.

Example **High MOI** single-cell CRISPR screen data

	Gene 1	Gene 2	...	Gene 5000
Cell 1	2	5		1
Cell 2	1	9		0
Cell 3	0	8		0
Cell 100,00	3	2		2

Gene matrix

	gRNA 1	gRNA 2	...	gRNA 500
Cell 1	TRUE	FALSE		FALSE
Cell 2	FALSE	FALSE		TRUE
Cell 3	TRUE	TRUE		TRUE
Cell 100,00	FALSE	TRUE		FALSE

gRNA matrix

	Conf 1	Conf 2	...	Conf 6
Cell 1				
Cell 2				
Cell 3				
Cell 100,00				

Confounder matrix

Example **High MOI** single-cell CRISPR screen data.

	Gene 1	Gene 2	...	Gene 5000
Cell 1	2	5		1
Cell 2	1	9		0
Cell 3	0	8		0
Cell 100,000	3	2		2

Gene matrix

	gRNA 1	gRNA 2	...	gRNA 500
Cell 1	TRUE	FALSE		FALSE
Cell 2	FALSE	FALSE		TRUE
Cell 3	TRUE	TRUE		TRUE
Cell 100,000	FALSE	TRUE		FALSE

gRNA matrix

	Conf 1	Conf 2	...	Conf 6
Cell 1				
Cell 2				
Cell 3				
Cell 100,000				

Confounder matrix

Proceed one gene and one gRNA at a time.

Example **High MOI** single-cell CRISPR screen data.

	Gene 1	Gene 2	...	Gene 5000
Cell 1	2	5		1
Cell 2	1	9		0
Cell 3	0	8		0
Cell 100,000	3	2		2

Gene matrix

	gRNA 1	gRNA 2	...	gRNA 500
Cell 1	TRUE	FALSE		FALSE
Cell 2	FALSE	FALSE		TRUE
Cell 3	TRUE	TRUE		TRUE
Cell 100,000	FALSE	TRUE		FALSE

gRNA matrix

	Conf 1	Conf 2	...	Conf 6
Cell 1				
Cell 2				
Cell 3				
Cell 100,000				

Confounder matrix

Proceed one gene and one gRNA at a time.

Example **High MOI** single-cell CRISPR screen data.

	Gene 1	Gene 2	...	Gene 5000
Cell 1	2	5		1
Cell 2	1	9		0
Cell 3	0	8		0
Cell 100,000	3	2		2

Gene matrix

	gRNA 1	gRNA 2	...	gRNA 500
Cell 1	TRUE	FALSE		FALSE
Cell 2	FALSE	FALSE		TRUE
Cell 3	TRUE	TRUE		TRUE
Cell 100,000	FALSE	TRUE		FALSE

gRNA matrix

	Conf 1	Conf 2	...	Conf 6
Cell 1				
Cell 2				
Cell 3				
Cell 100,000				

Confounder matrix

Proceed one gene and one gRNA at a time.

SCEPTRE is based on the **conditional randomization test (CRT)**, overcoming limitations of previous parametric and nonparametric approaches.

	Adjusts for confounders	Is robust to expression model misspecification
Parametric method	✅	❌
Nonparametric method	❌	✅
Conditional randomization test	✅	✅

SCEPTRE is a custom implementation of the CRT for single-cell CRISPR screen data.

SCEPTRE

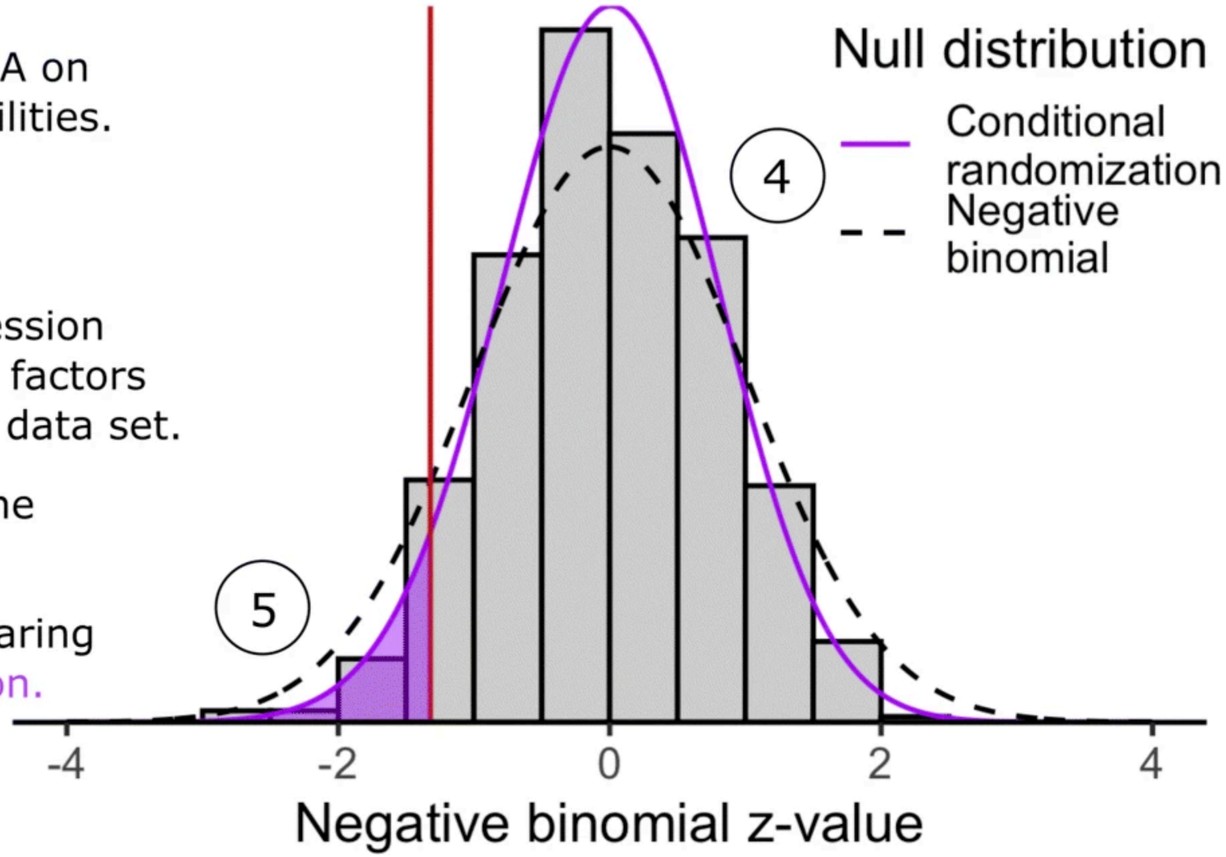
Step 1: Fit logistic regression of gRNA on technical factors to get fitted probabilities.

Step 2: Repeatedly resample gRNAs for each cell based on probabilities.

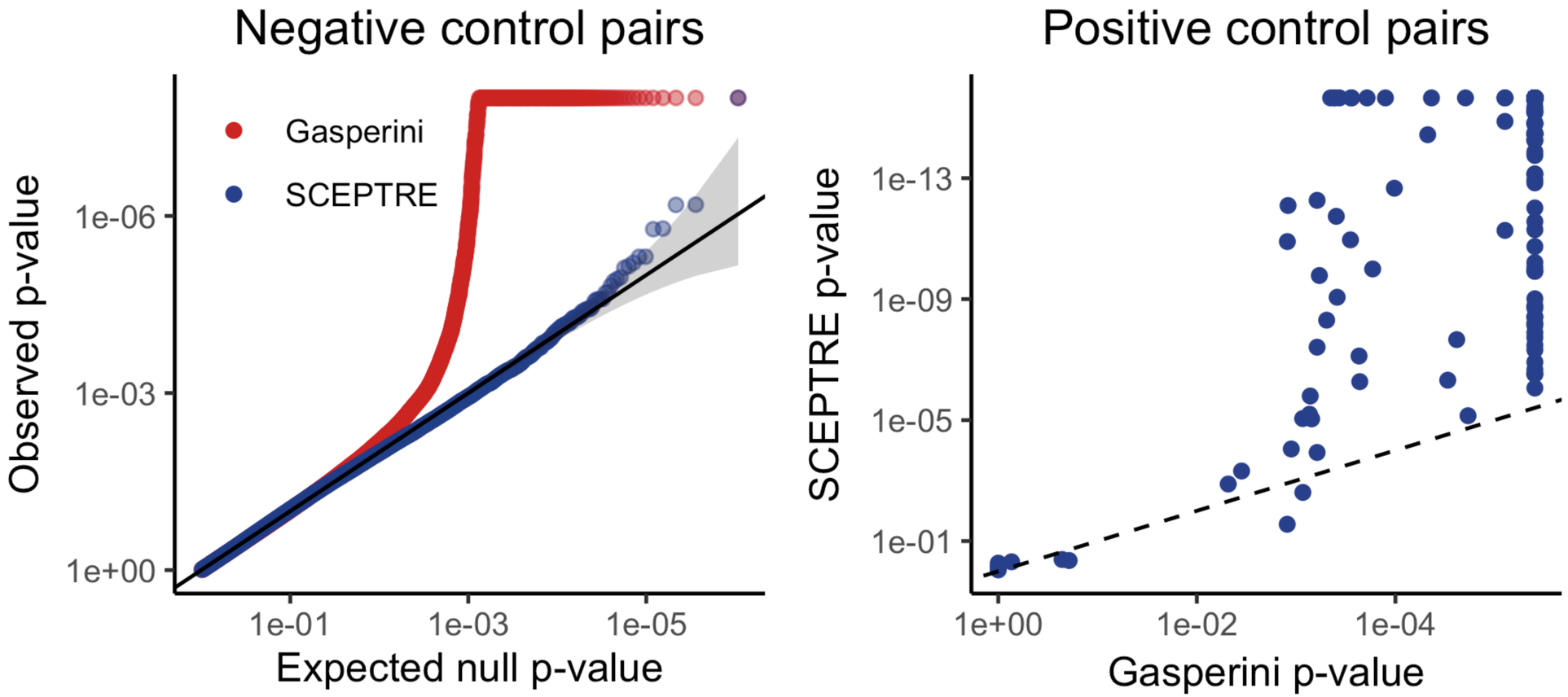
Step 3: Fit a negative binomial regression of expression on gRNA and technical factors for original data and each reshuffled data set.

Step 4: Fit a skew-t distribution to the set of resampled z-values.

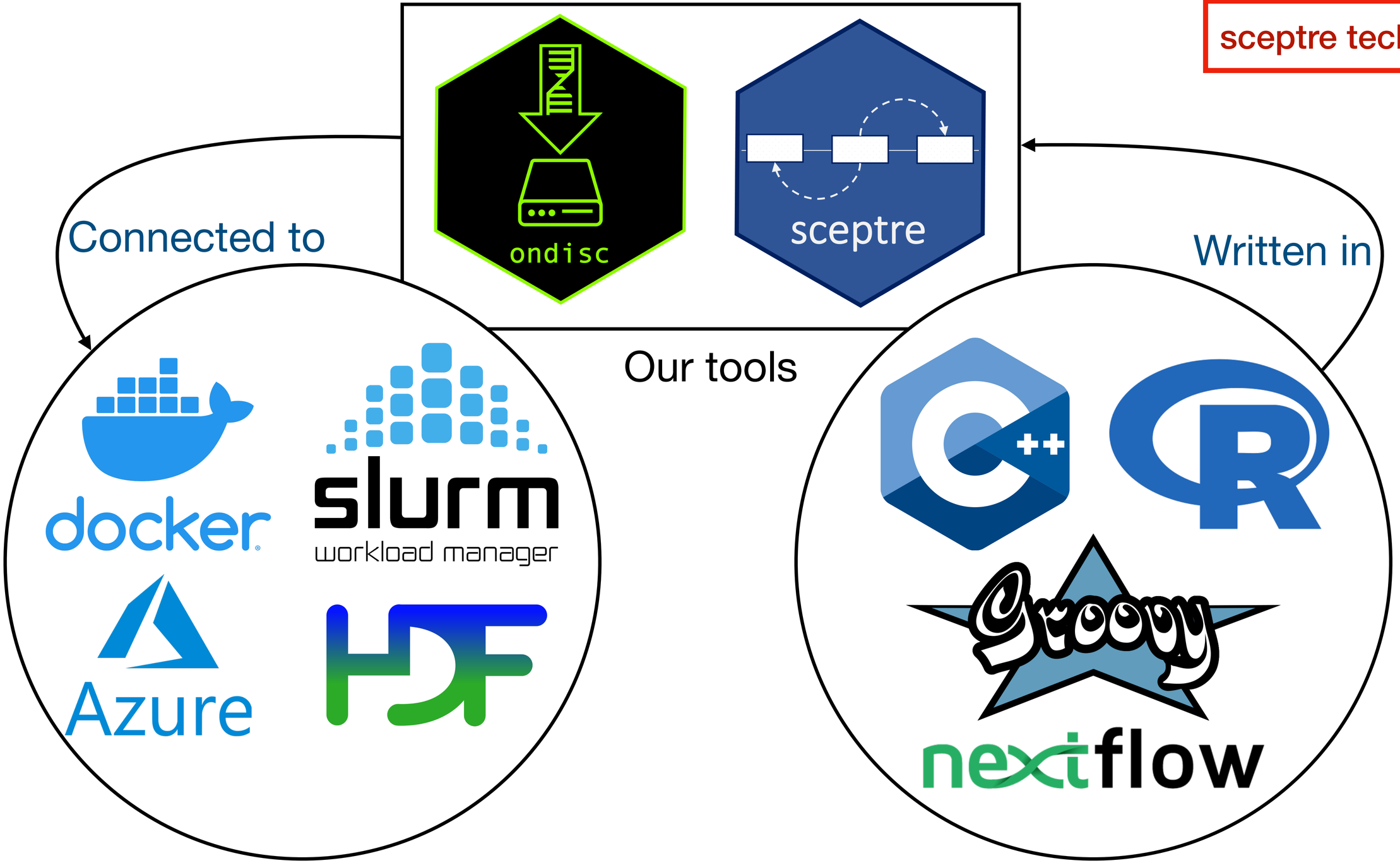
Step 5: Compute a p-value by comparing original z-value to the null distribution.



SCEPTRE demonstrates superior calibration and power on negative and positive control data, respectively.



sceptre tech stack



The sceptre technology stack is in use by several research groups worldwide.

ICR The Institute of Cancer Research

(UK; cancer)

NEW YORK GENOME CENTER®

(USA; blood diseases)

Calico

(USA; lifespan extension)

MONASH University

(Australia; cancer)

City of Hope®

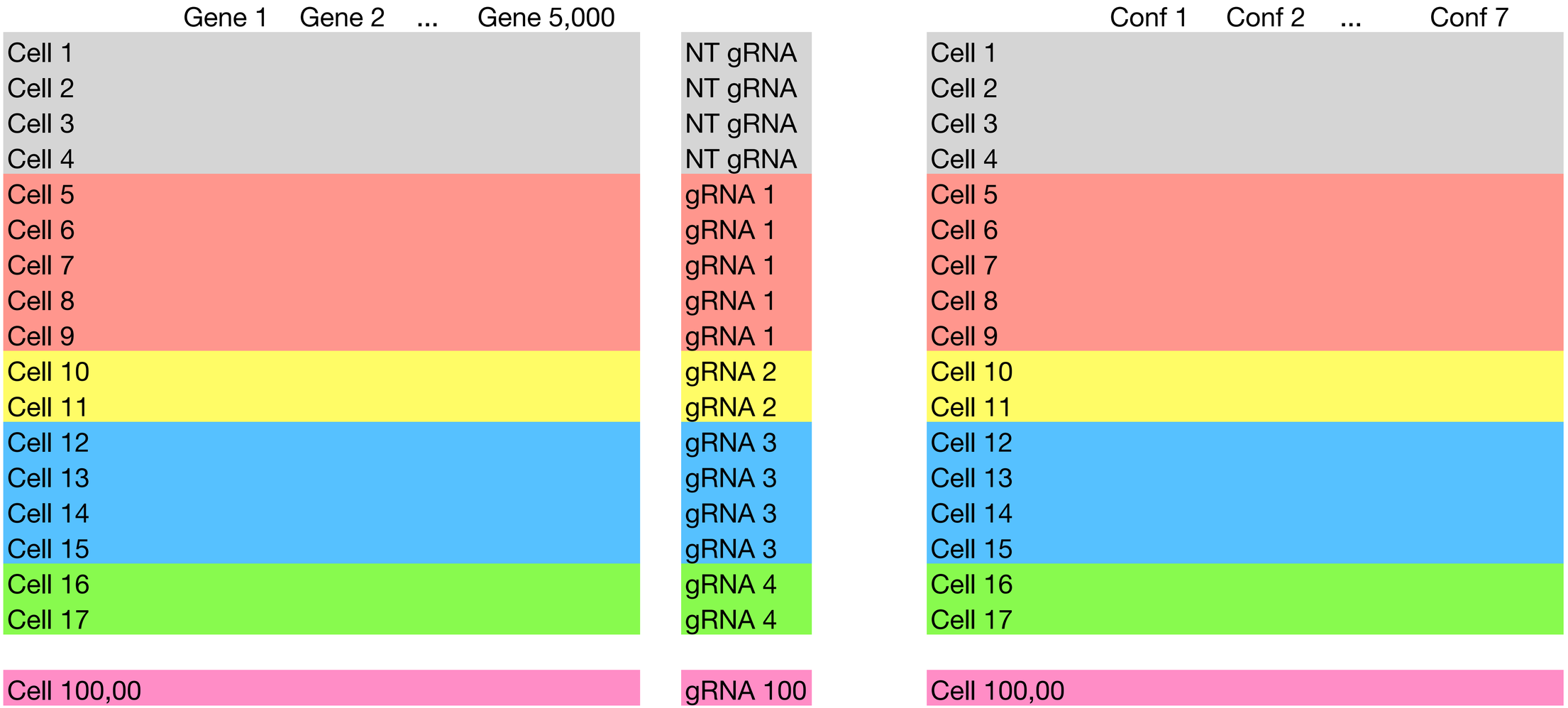
(USA)

Roadmap

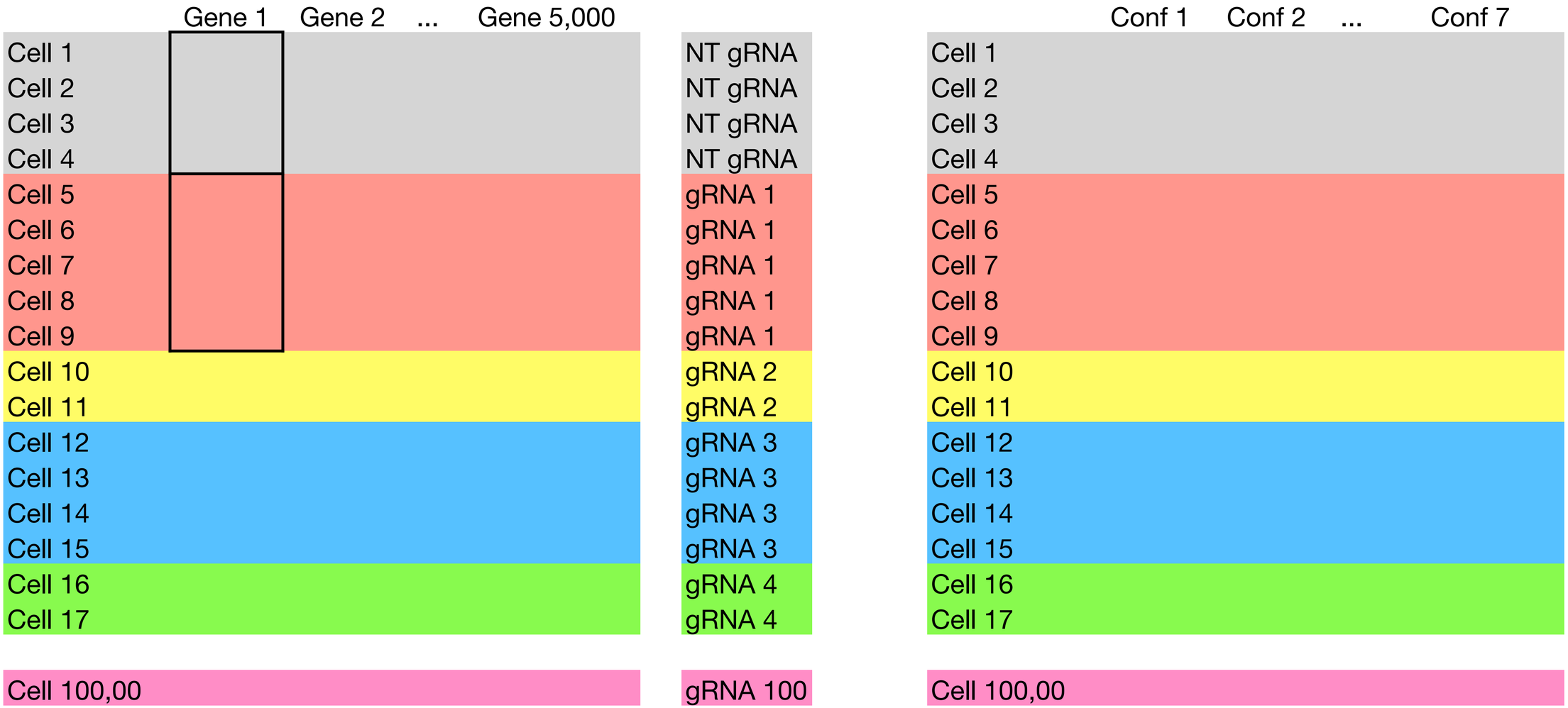
1. SCEPTRE in high MOI

2. SCEPTRE in low MOI

Low MOI datasets involve inserting a **single** gRNA into each cell.

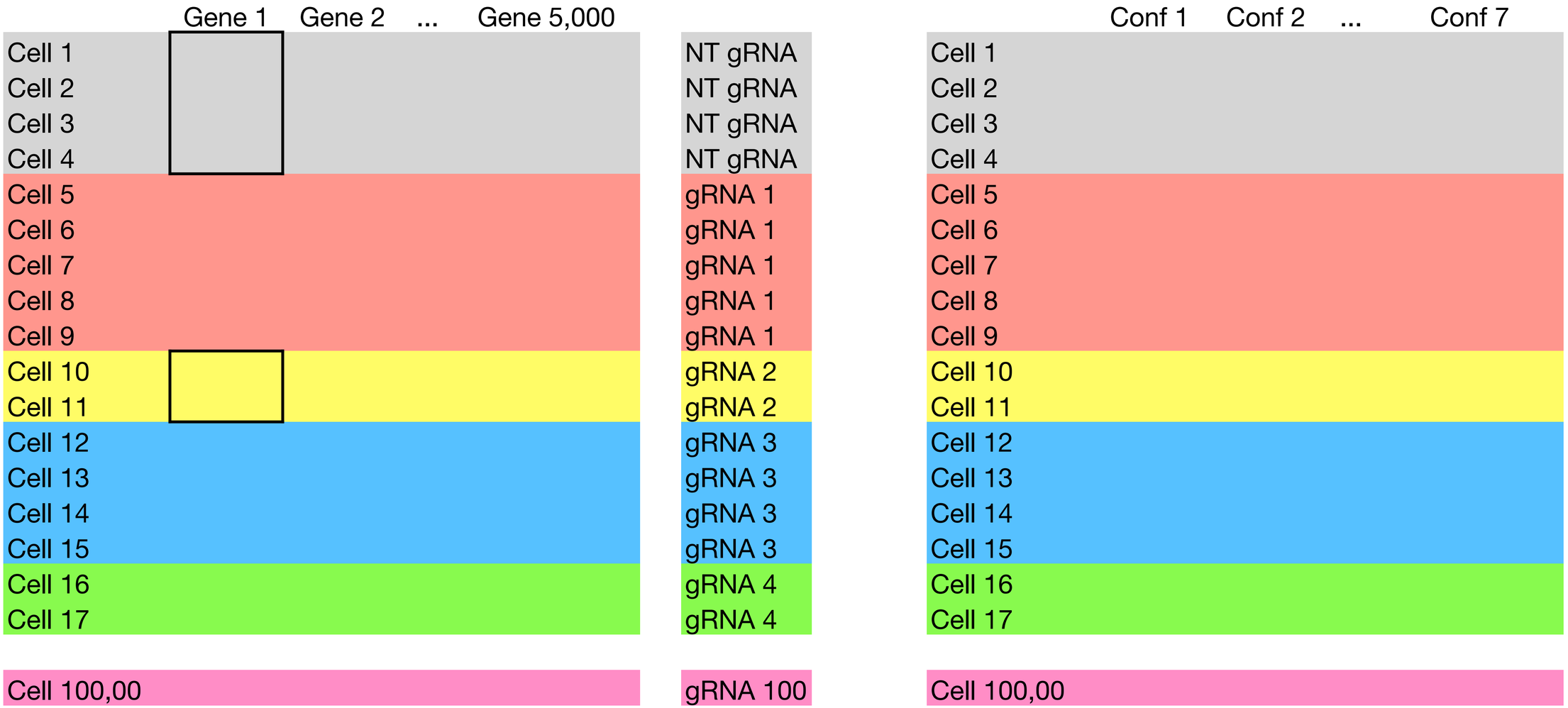


Proceed one gene and one gRNA at a time, comparing "treatment" cells to NT cells.



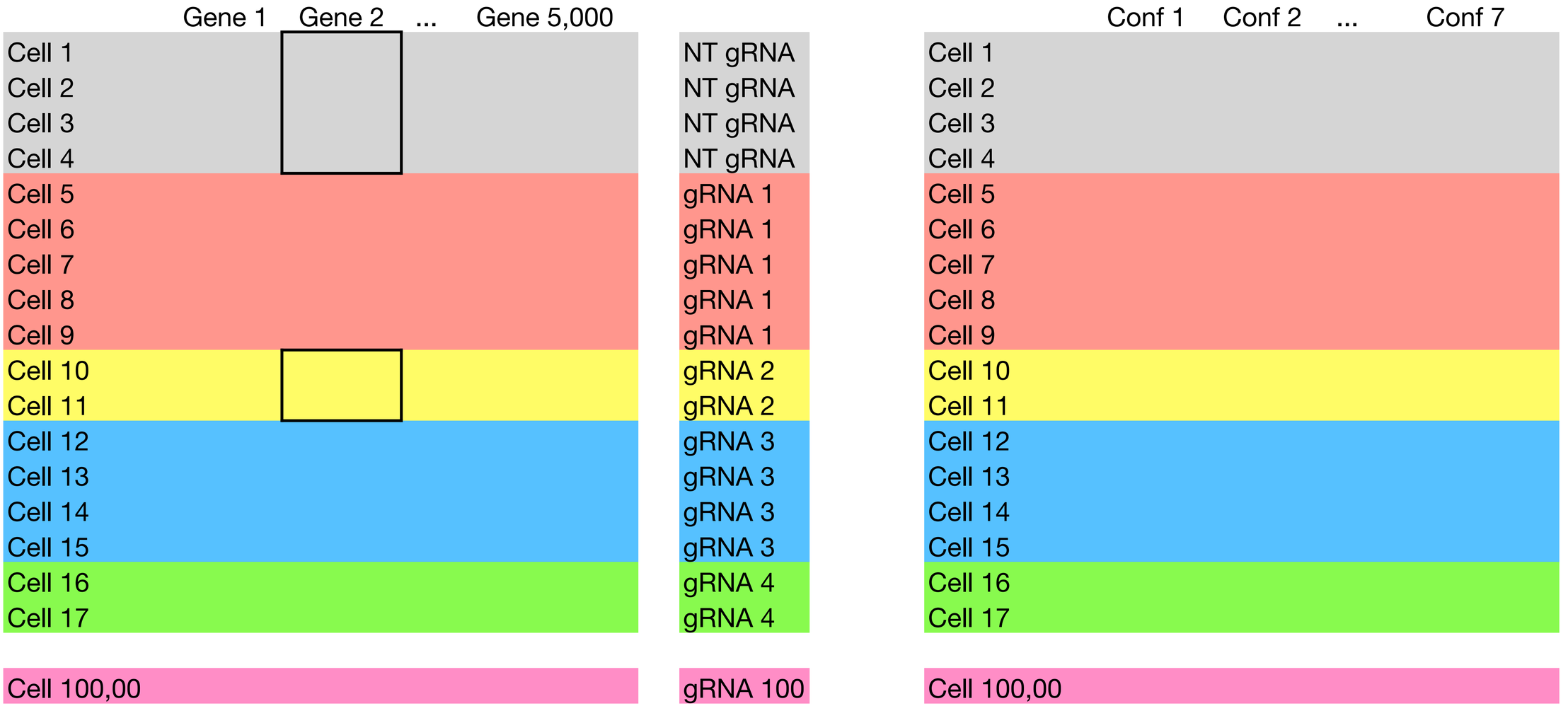
Gene matrix **gRNA vector** **Covariate matrix**

Proceed one gene and one gRNA at a time, comparing "treatment" cells to NT cells.



Gene matrix **gRNA vector** **Covariate matrix**

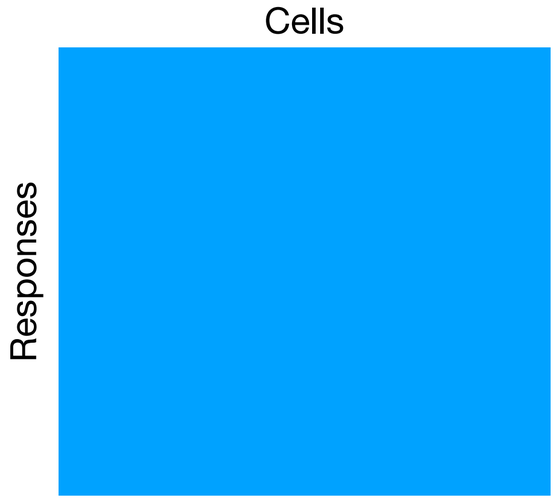
Proceed one gene and one gRNA at a time, comparing "treatment" cells to NT cells.



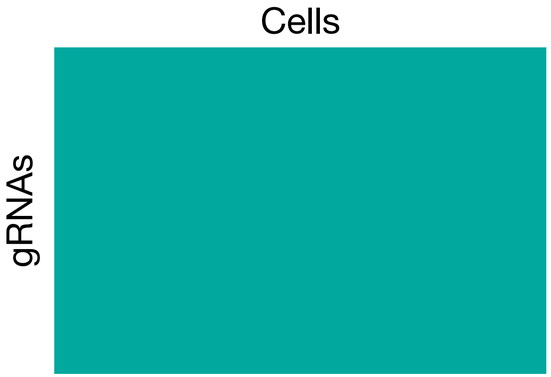
Gene matrix **gRNA vector** **Covariate matrix**

Undercover gRNA calibration assessment

- A single-cell CRISPR screen differential expression method takes as input four arguments (T, targeting; NT, non-targeting):



1. Response-by-cell expression matrix



2. gRNA-by-cell expression matrix

ID	Type
gRNA-1	T
gRNA-2	T
gRNA-3	T
gRNA-4	T
gRNA-5	NT
gRNA-6	NT
gRNA-7	NT
gRNA-8	NT

3. gRNA information table

4. Cell-specific covariates

Undercover gRNA calibration assessment

ID	Group	Type
gRNA-1	group-1	T
gRNA-2	group-1	T
gRNA-3	group-2	T
gRNA-4	group-2	T
gRNA-5	.	NT
gRNA-6	.	NT
gRNA-7	.	NT
gRNA-8	.	NT

True gRNA information table

Label swap

ID	Type
gRNA-1	T
gRNA-2	T
gRNA-3	T
gRNA-4	T
gRNA-5	NT
gRNA-6	T
gRNA-7	NT
gRNA-8	NT

Swapped gRNA information table

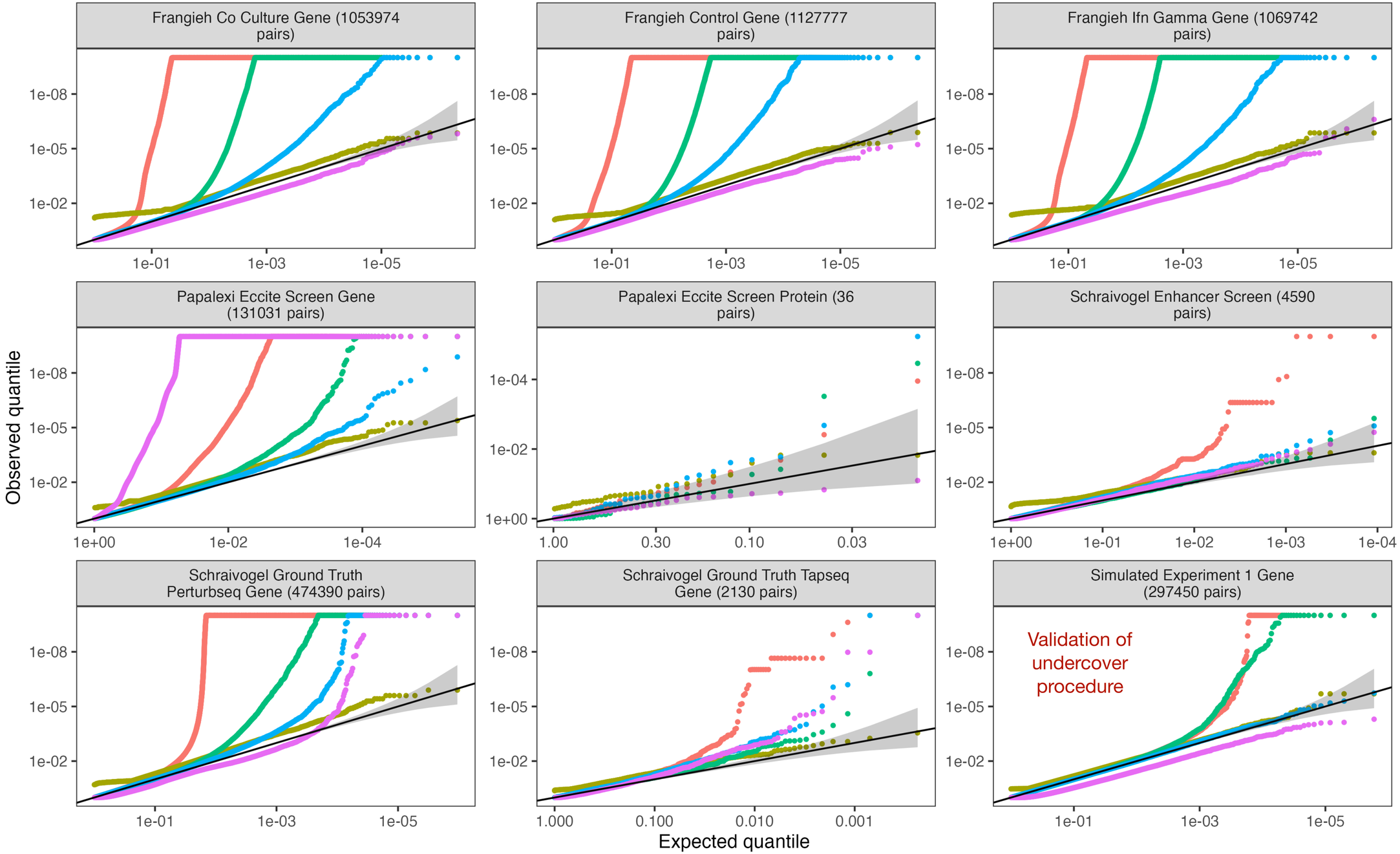
- Keep the gene expression, gRNA expression, and cell covariate matrices the same. Test for differential expression between the undercover gRNA and all genes.

We leveraged the undercover gRNA framework to benchmark the calibration of five methods on 10+ real datasets.

- 10 real datasets published across four papers, plus one simulated dataset.
  - **Frangieh, Nature Genetics 2021 (3)**
  - **Papalexi, Nature Genetics, 2021 (2)**
  - **Schraivogel, Nature Methods, 2020 (4)**
  - **Liscovitch, Nature Biotechnology, 2021 (2)**
  - **Simulated dataset**

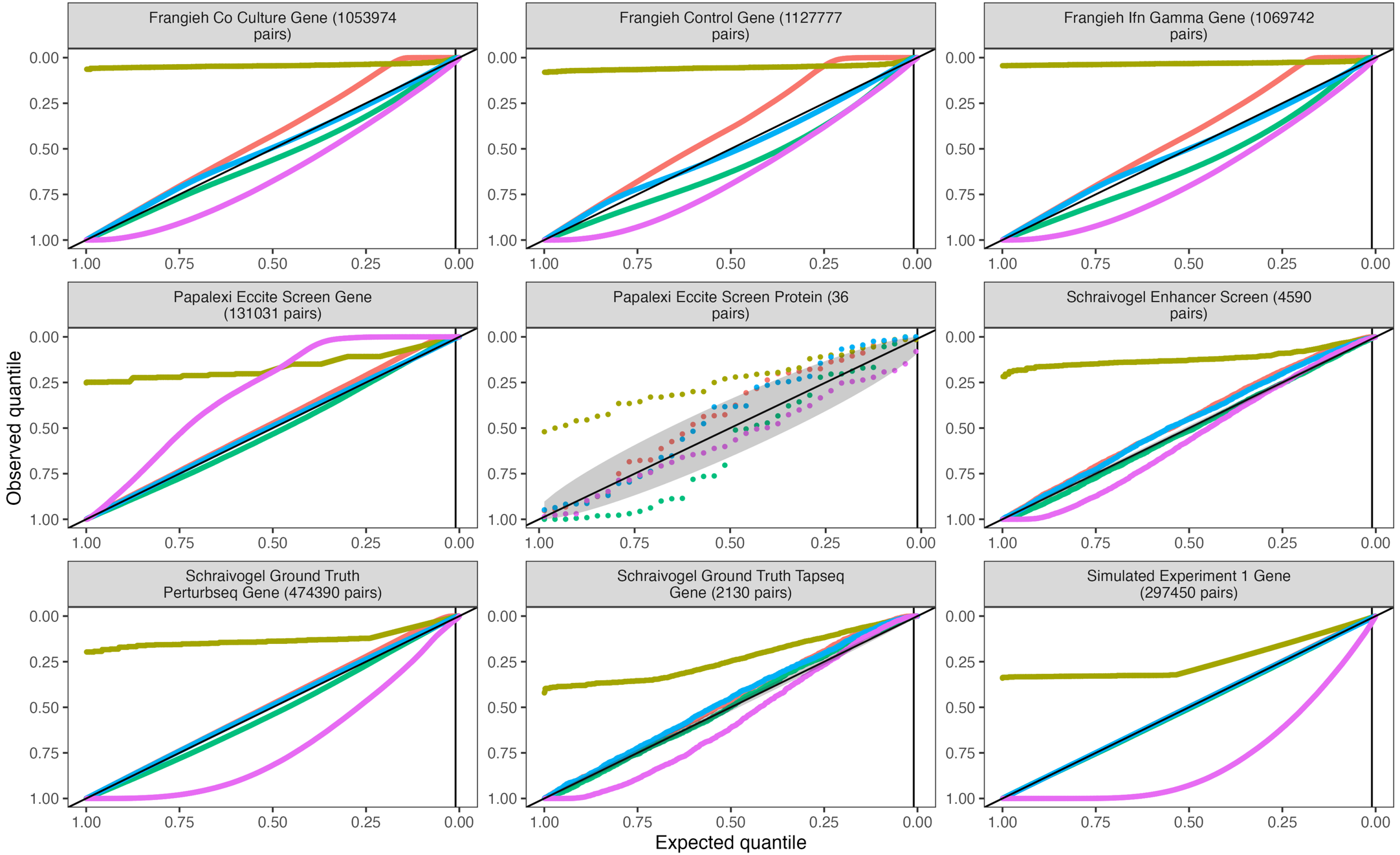
Existing methods (normalizing the data, carrying out the differential expression test)

- 1. **MIMOSCA**. Permutation test with ridge test stat (used by Frangieh)
- 2. **Seurat DE**. Mann-Whitney test (used by Papalexi)
- 3. **Schraivogel method**. Hurdle model (used by Schraivogel)
- 4. **Liscovitch method**. T-test (used by Liscovitch)
- 5. **Weissman Method**. KS test (used by the Weissman lab)



Transformed QQ-plot

Method · Liscovitch Method · Mimosca · Schraivogel Method · Seurat De · Weissman Method

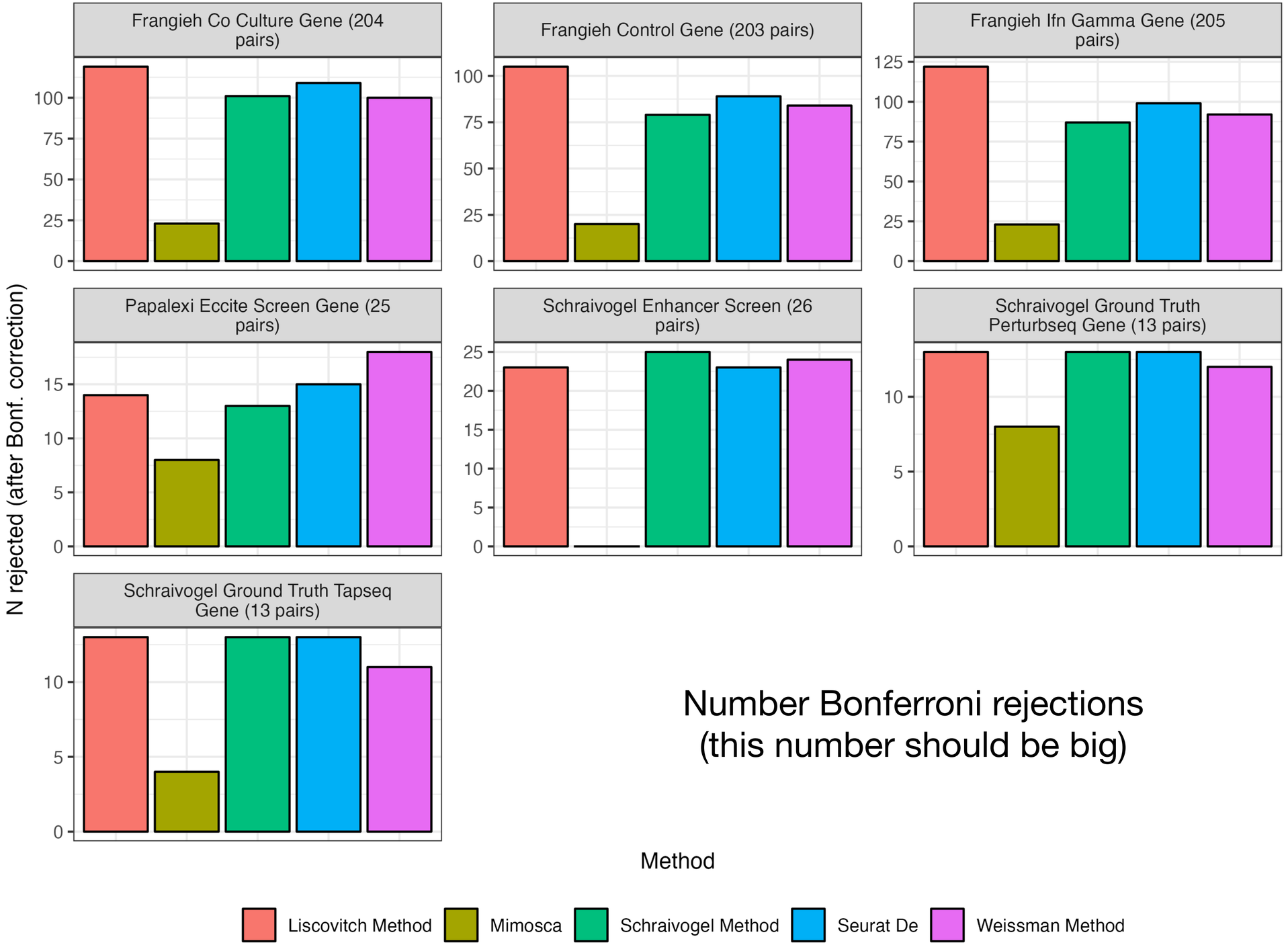

Frangieh Co Culture Gene (1053974 pairs)
Frangieh Control Gene (1127777 pairs)
Frangieh Ifn Gamma Gene (1069742 pairs)
Papalexi Eccite Screen Gene (131031 pairs)
Papalexi Eccite Screen Protein (36 pairs)
Schraivogel Enhancer Screen (4590 pairs)
Schraivogel Ground Truth Perturbseq Gene (474390 pairs)
Schraivogel Ground Truth Tapseq Gene (2130 pairs)
Simulated Experiment 1 Gene (297450 pairs)
Observed quantile
Expected quantile
0.00
0.25
0.50
0.75
1.00
Method
Liscovitch Method
Mimosca
Schraivogel Method
Seurat De
Weissman Method

Raw QQ-plot


Frangieh Co Culture Gene (1053974 pairs)
Frangieh Control Gene (1127777 pairs)
Frangieh Ifn Gamma Gene (1069742 pairs)
Papalexi Eccite Screen Gene (131031 pairs)
Papalexi Eccite Screen Protein (36 pairs)
Schraivogel Enhancer Screen (4590 pairs)
Schraivogel Ground Truth Perturbseq Gene (474390 pairs)
Schraivogel Ground Truth Tapseq Gene (2130 pairs)
Simulated Experiment 1 Gene (297450 pairs)
1e+05
1e+03
1e+01
N rejected (after Bonf. correction)
Method
Number Bonferroni rejections
Liscovitch Method
Mimosca
Schraivogel Method
Seurat De
Weissman Method

Positive control gRNA power assessment

- Pair gRNAs that target gene transcription start sites or known enhancers to the genes targeted by these elements.

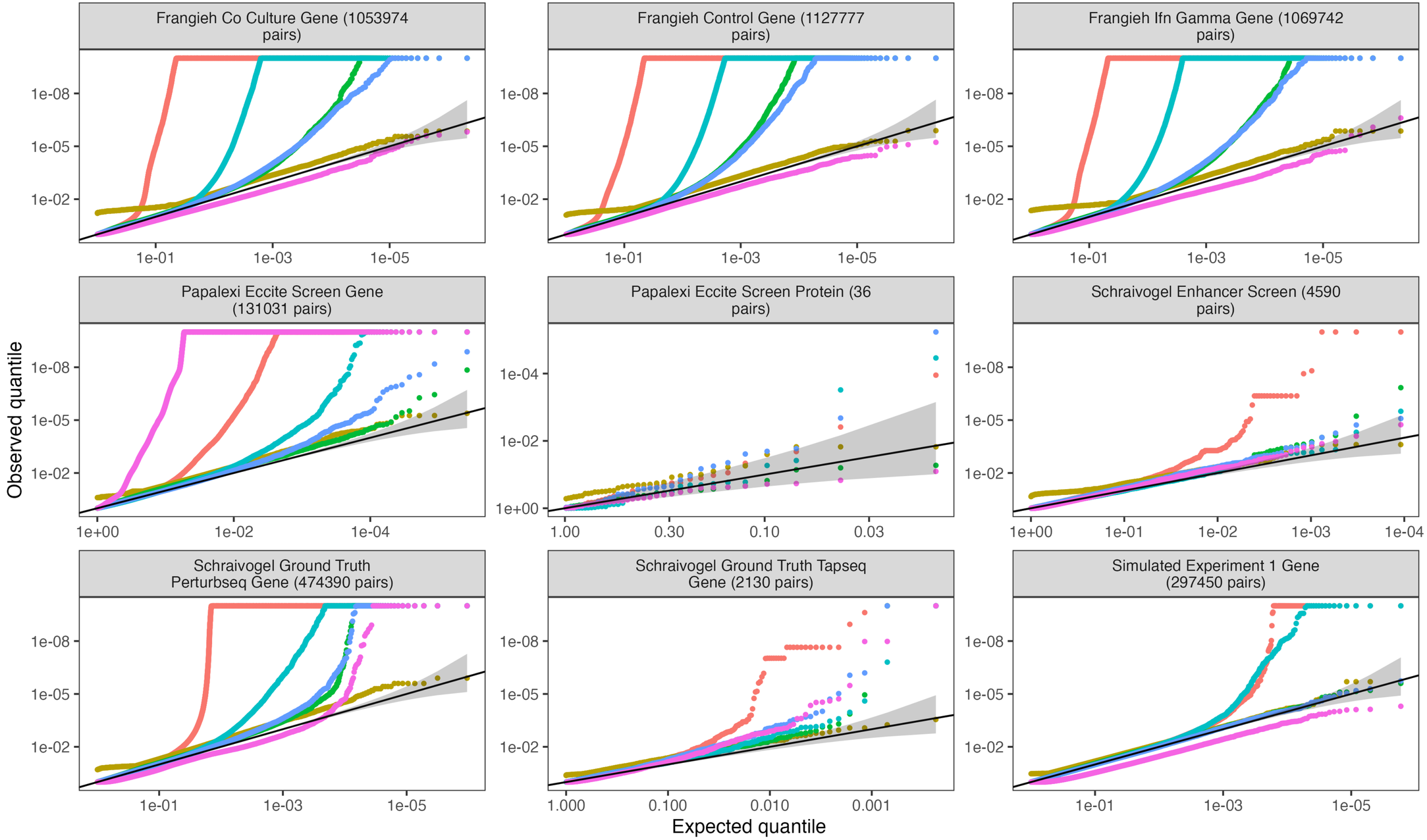

Frangieh Co Culture Gene (204 pairs)
Frangieh Control Gene (203 pairs)
Frangieh Ifn Gamma Gene (205 pairs)
Papalexi Eccite Screen Gene (25 pairs)
Schraivogel Enhancer Screen (26 pairs)
Schraivogel Ground Truth Perturbseq Gene (13 pairs)
Schraivogel Ground Truth Tapseq Gene (13 pairs)
N rejected (after Bonf. correction)
Number Bonferroni rejections
(this number should be big)
Method
Liscovitch Method
Mimosca
Schraivogel Method
Seurat De
Weissman Method

No method that we examined exhibits good calibration *and* good power.

Method	Calibration	Power
MIMOSCA	Tail calibration OK, bulk calibration poor	✗
Seurat	✗	OK
Schraivogel method	✗	OK
Liscovitch method	✗	OK
Weissman method	✗	OK

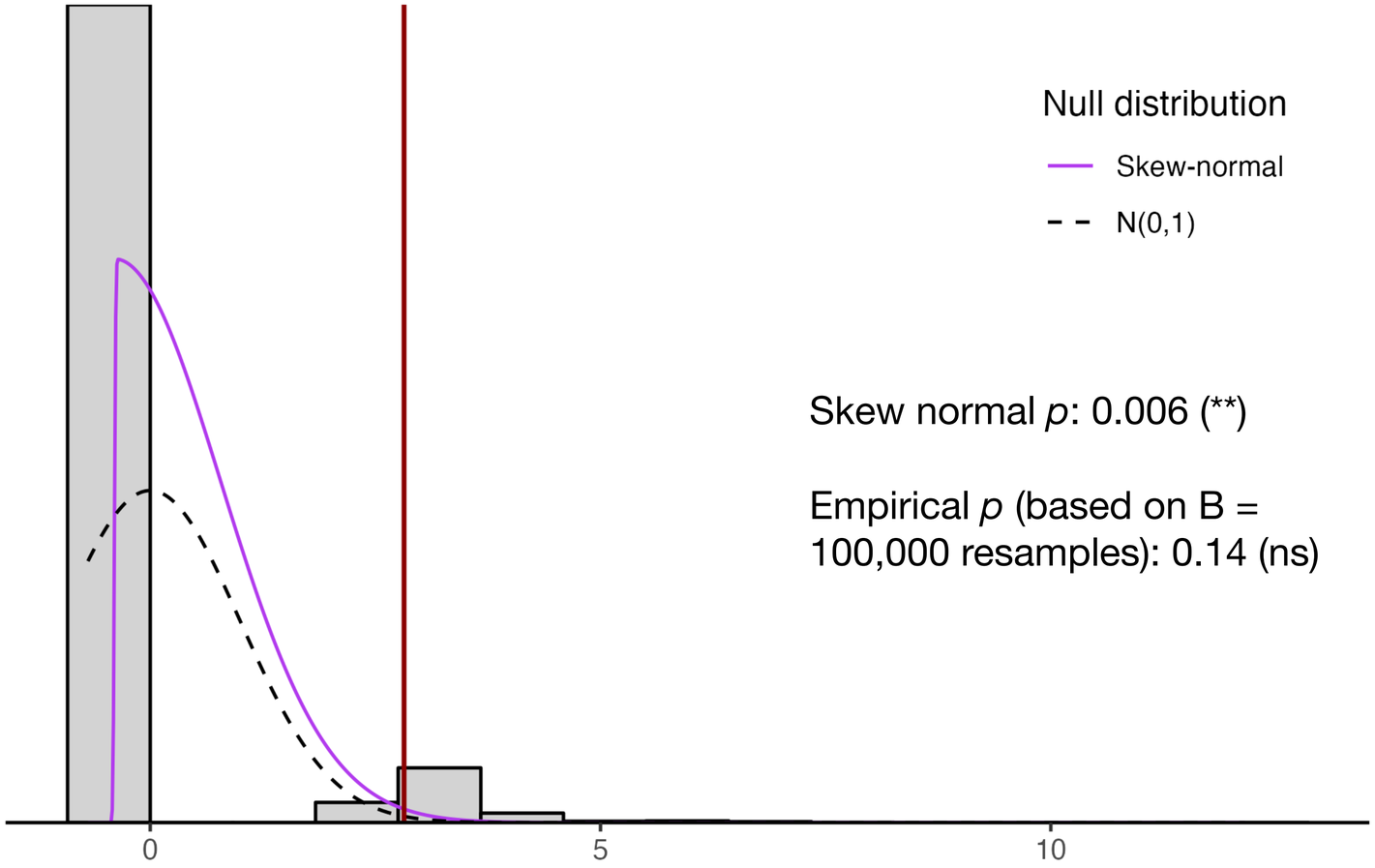
sceptre in low MOI: three basic ingredients

1. A resampling mechanism (**permutation** or conditional resampling)
2. A test statistic (negative binomial score statistic)
3. A procedure for reducing the number of resamples (**fit skew normal**)



sceptre currently exhibits inflation :(

Probable primary culprit: poor skew-normal fits to resampling distributions.



Current objective: solve this problem!

Either (i) avoid parametric approximations altogether, or (ii) verify the correctness of parametric assumptions using a data-dependent procedure.

Why I like working on single-cell CRISPR screen applications

1. Potential to make a fairly large scientific splash (quixotic as it sounds).

2. Develop statistical perspectives and ideas that one might not have otherwise developed.

    - Many interesting issues: confounding, model misspecification, discreteness, availability of negative/ positive controls, computational constraints, etc.