

Bioinfo single-cell seminar

29-11-2022

Perturb-seq analyses and challenges

Paul KLEIN (bioinformatician)

Joshua WATERFALL group

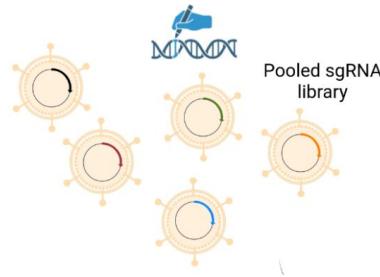
Integrative Functional Genomics of Cancer (IFGC)

Unit U830, Translational Research Dpt

Perturb-seq basic principles

Perturb-seq enables to perform single cell RNA sequencing (**scRNA-seq**) on pooled genetic perturbation screens (**CRISPR**).

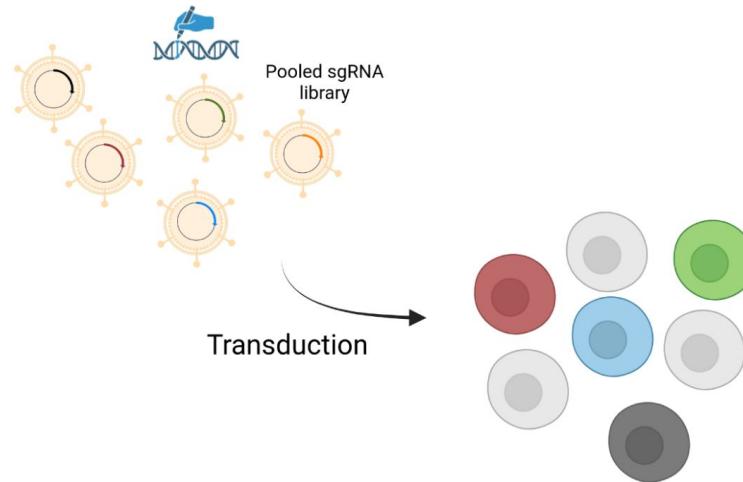
GOAL = infer gene function by studying the transcriptomic phenotypes when the gene is knocked down (reverse genetics).



Perturb-seq basic principles

Perturb-seq enables to perform single cell RNA sequencing (**scRNA-seq**) on pooled genetic perturbation screens (**CRISPR**).

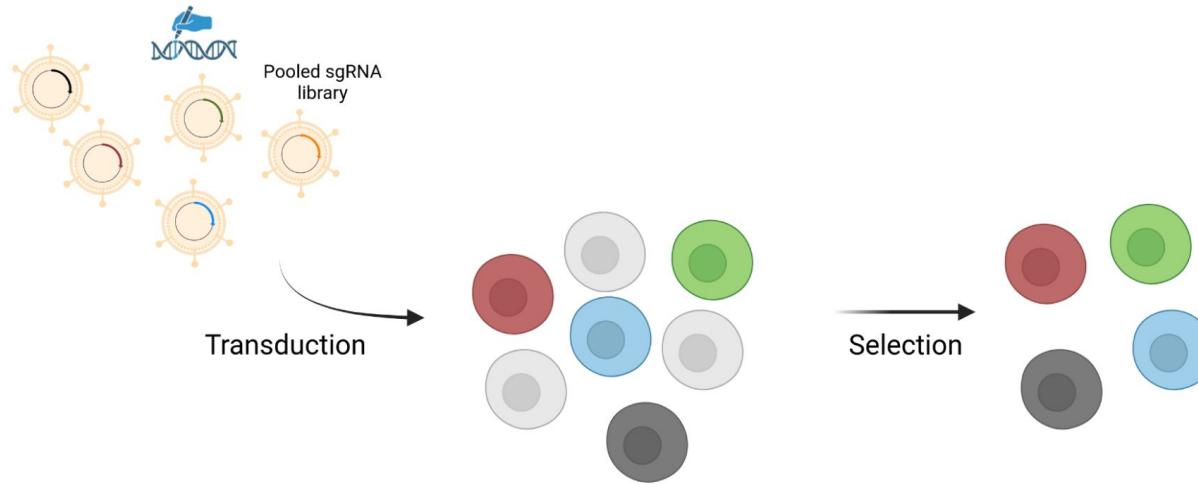
GOAL = infer gene function by studying the transcriptomic phenotypes when the gene is knocked down (reverse genetics).



Perturb-seq basic principles

Perturb-seq enables to perform single cell RNA sequencing (**scRNA-seq**) on pooled genetic perturbation screens (**CRISPR**).

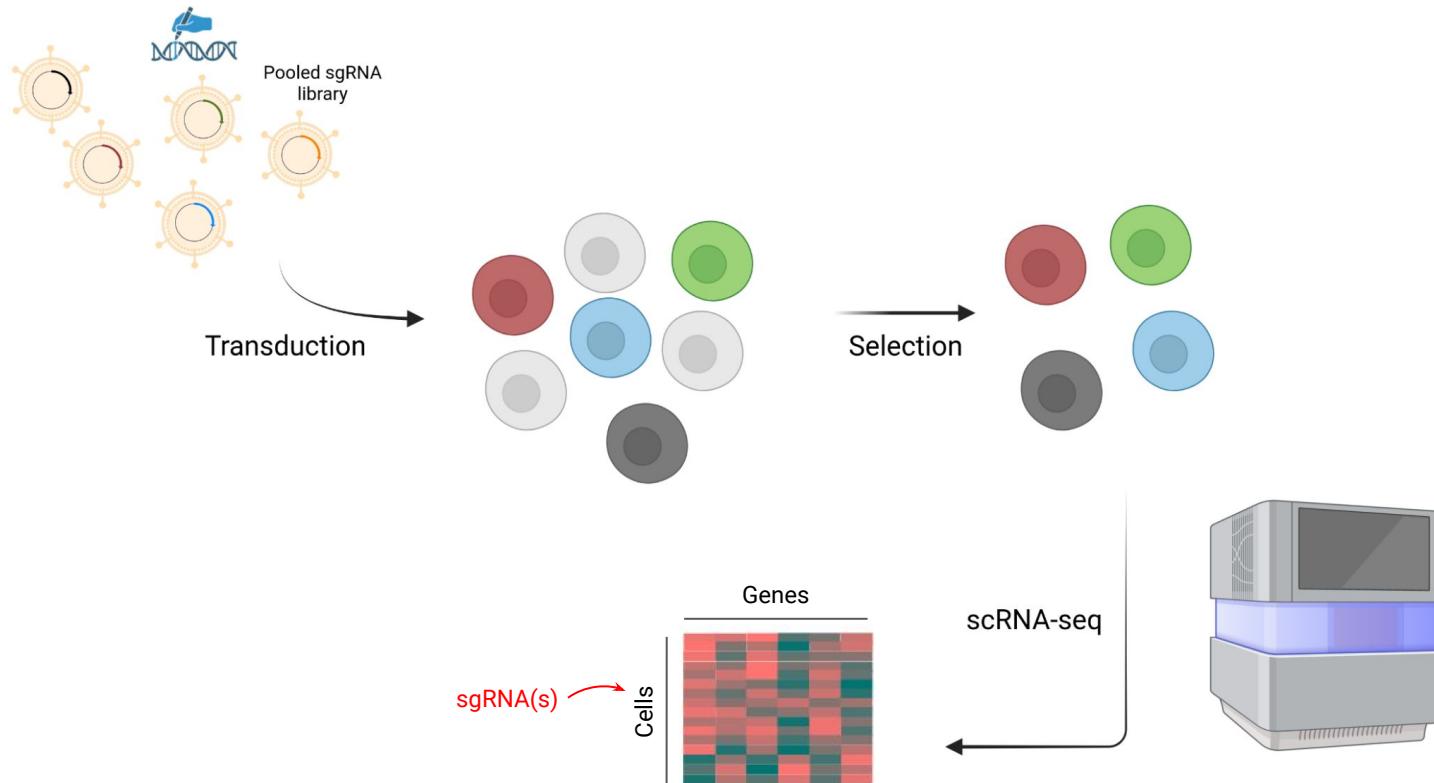
GOAL = infer gene function by studying the transcriptomic phenotypes when the gene is knocked down (reverse genetics).



Perturb-seq basic principles

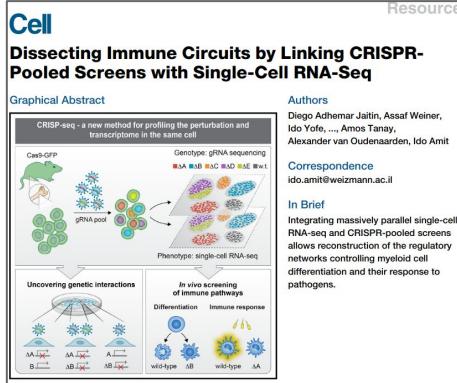
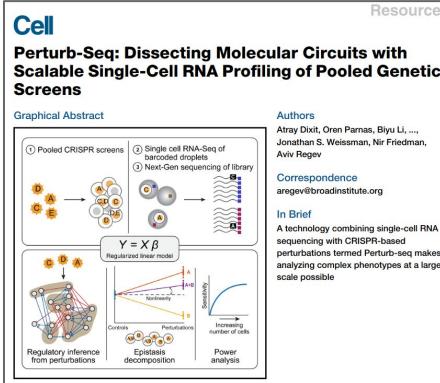
Perturb-seq enables to perform single cell RNA sequencing (**scRNA-seq**) on pooled genetic perturbation screens (**CRISPR**).

GOAL = infer gene function by studying the transcriptomic phenotypes when the gene is knocked down (reverse genetics).



Combining CRISPR perturbations with scRNA-seq

Perturb-seq



2016



nature methods

Explore content ▾ About the journal ▾ Publish with us ▾ Subscribe

nature > nature methods > articles > article

Published: 18 January 2017

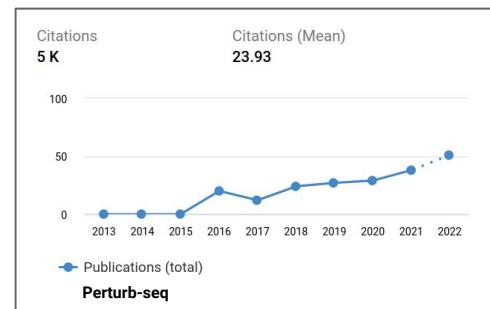
Pooled CRISPR screening with single-cell transcriptome readout

Paul Datlinger, André F Rendeiro, Christian Schmidl, Thomas Krausgruber, Peter Traxler, Johanna Klughammer, Linda Schuster, Amelie Kuchler, Donat Alpar & Christoph Bock

Nature Methods 14, 297–301 (2017) | Cite this article

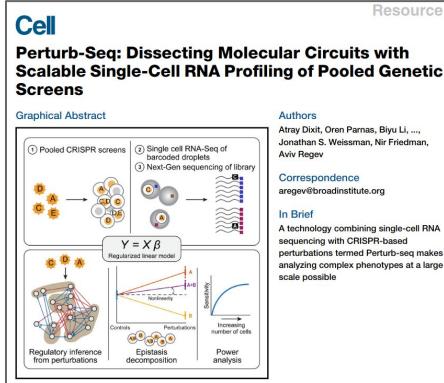
70k Accesses | 405 Citations | 151 Altmetric | Metrics

CROP-seq

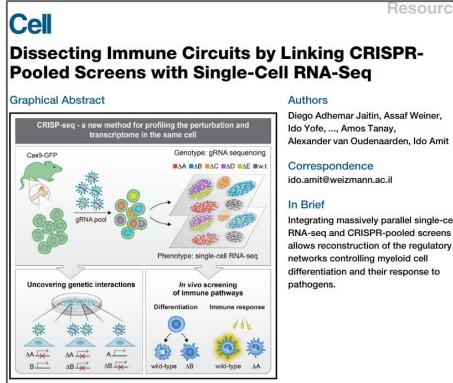


Combining CRISPR perturbations with scRNA-seq

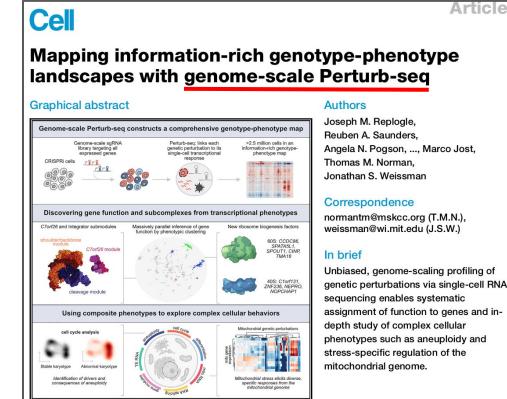
Perturb-seq



CRISP-seq



2016



2022

nature methods

Explore content ▾ About the journal ▾ Publish with us ▾ Subscribe

nature > nature methods > articles > article

Published: 18 January 2017

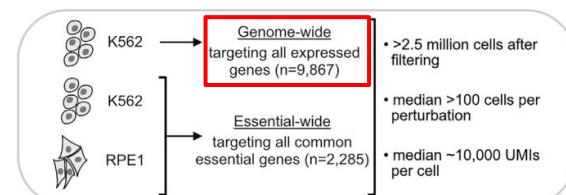
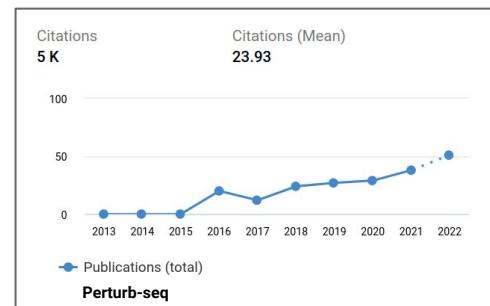
Pooled CRISPR screening with single-cell transcriptome readout

Paul Datlinger, André F Rendeiro, Christian Schmidl, Thomas Krausgruber, Peter Traxler, Johanna Klughammer, Linda Schuster, Amelie Kuchler, Donat Alpar & Christoph Bock

Nature Methods 14, 297–301 (2017) | Cite this article

70k Accesses | 405 Citations | 151 Altmetric | Metrics

CROP-seq



Preliminary dataset

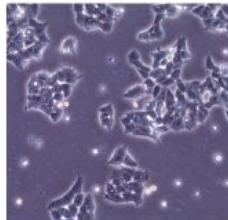
Mechanisms of Repression by Polycomb Group Proteins (UMR3215 / U934)



Raphaël
MARGUERON



Michel
WASLEF



HAP1 cells

1

Experimental details

- 340 genes targeted (chromatin regulators)
- 3 different sgRNA per target gene
- 30 Control sgRNA
- MOI=0.1 and MOI=3

1050 sgRNAs

2

Count Matrix generation

- cellranger (v7.0)

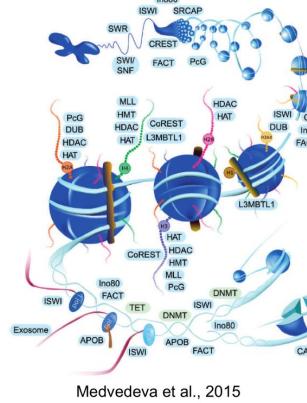
3

Pre-processing

- >1e3 genes per cell
- <10% mitochondrial reads

MOI=0.1	MOI=3
5,726	5,202

Number of cells after UMI
filtering and pre-processing



DNA modification
- DNA methylation
- DNA Hydroxymethylation

+ *DNA methylation readers*

Histone modification
- methylation
- acetylation
- phosphorylation
- ubiquitination
- sumoylation
- GlnAcylation
- citrullination

+ *readers*
+ *erasers (HDACs, demethylases, de-ubiquitinases, etc)*

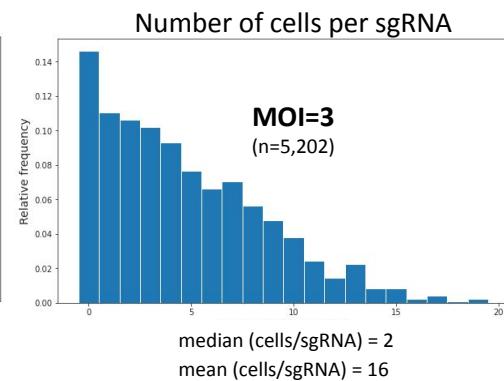
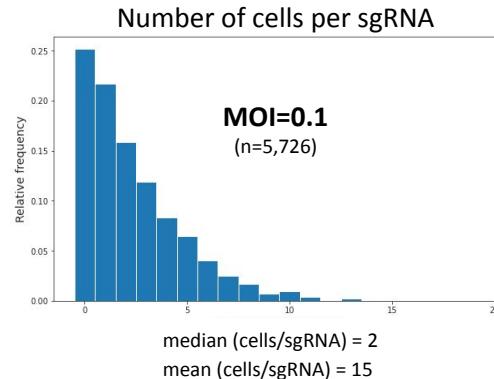
Chromatin remodeling

Histone chaperones

Which sgRNA induced an efficient target gene silencing ?

To assess sgRNA silencing efficiency, we need:

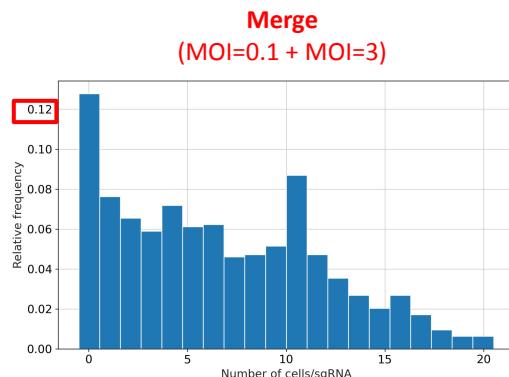
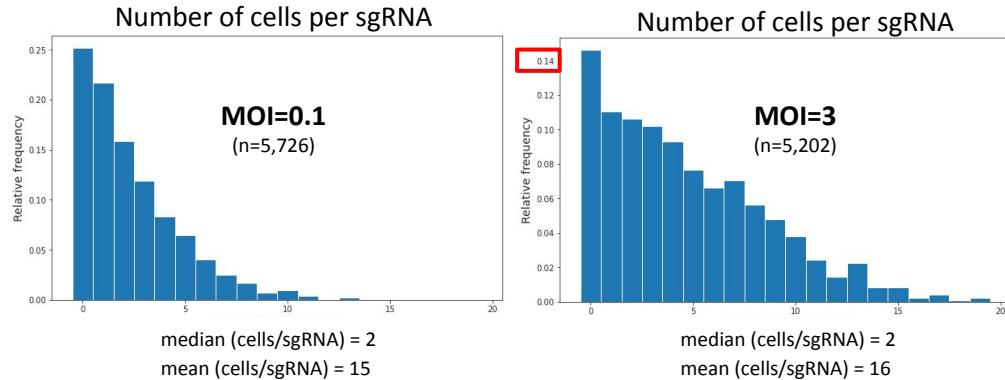
- a sufficient number of cells being infected by the sgRNA



Which sgRNA induced an efficient target gene silencing ?

To assess sgRNA silencing efficiency, we need:

- a sufficient number of cells being infected by the sgRNA

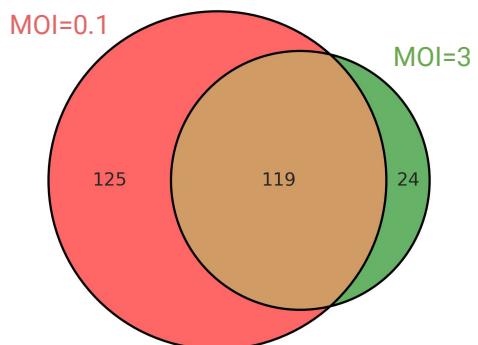


Which sgRNA induced an efficient target gene silencing ?

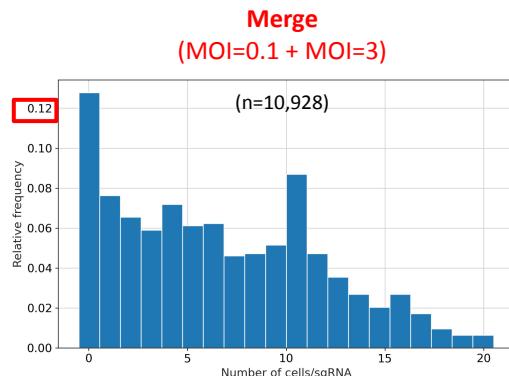
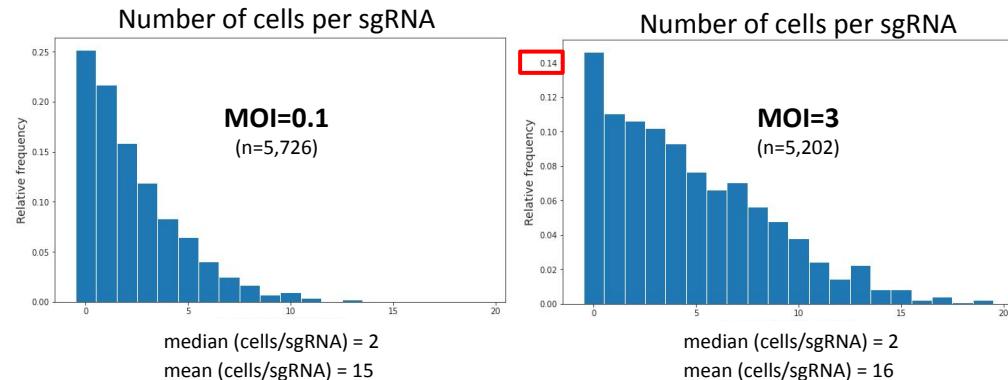
To assess sgRNA silencing efficiency, we need:

- a sufficient number of cells being infected by the sgRNA

sgRNAs that did not infect cells
(total sgRNA = 1,050)



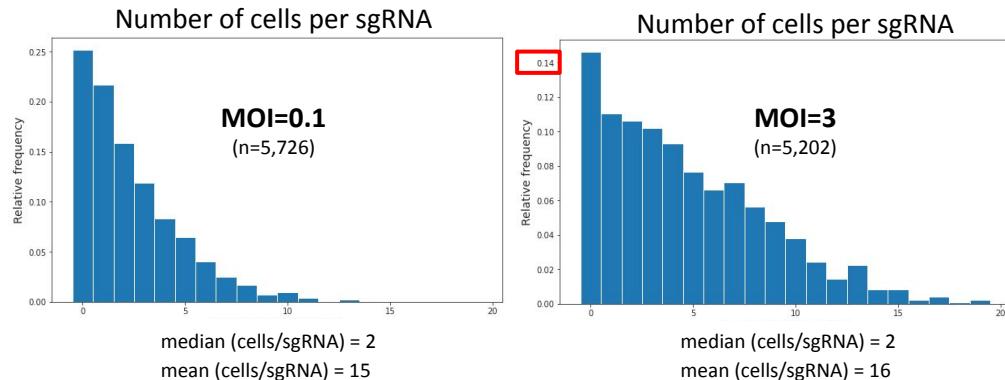
(Most) sgRNA that did not infect any cell with MOI=3, did not infect cells with MOI=0.1 neither.



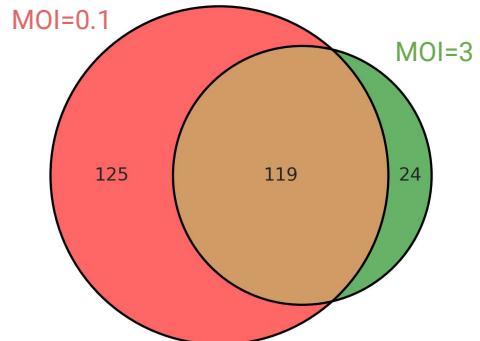
Which sgRNA induced an efficient target gene silencing ?

To assess sgRNA silencing efficiency, we need:

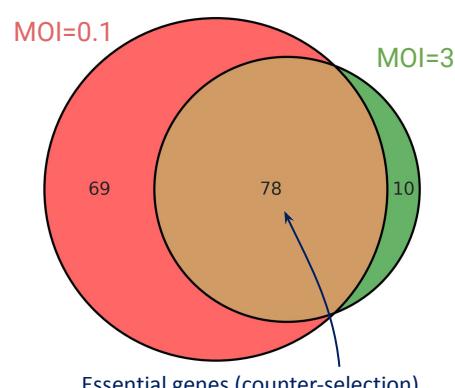
- a sufficient number of cells being infected by the sgRNA



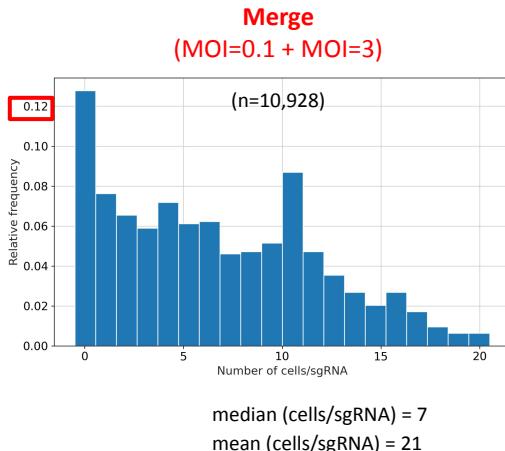
sgRNAs that did not infect cells
(total sgRNA = 1,050)



targeted genes with no infected cells
(total targeted genes = 340)



(Most) sgRNA that did not infect any cell with
MOI=3, did not infect cells with MOI=0.1 neither.



Which sgRNA induced an efficient target gene silencing ?

To assess sgRNA silencing efficiency, we need:

- a sufficient number of cells being infected by the sgRNA
- **being confident in the sgRNA assignment**

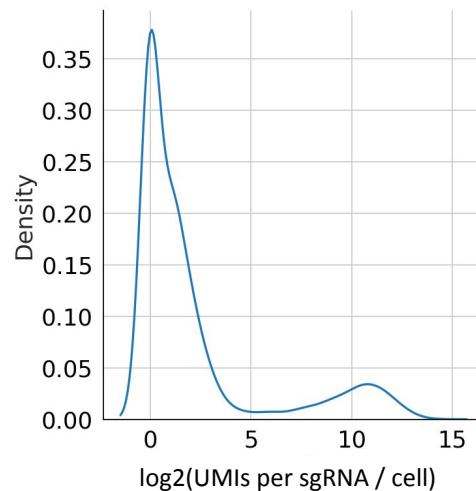
How sgRNA are assigned to cells ?

How to distinguish true signal from background protospacer UMIs ?

Which sgRNA induced an efficient target gene silencing ?

To assess sgRNA silencing efficiency, we need:

- a sufficient number of cells being infected by the sgRNA
- **being confident in the sgRNA assignment**



How sgRNA are assigned to cells ?

How to distinguish true signal from background protospacer UMIs ?

Cellranger sgRNA assignment call

UMI threshold is automatically computed by cellranger.

It is done per sgRNA, independent of all other sgRNAs.

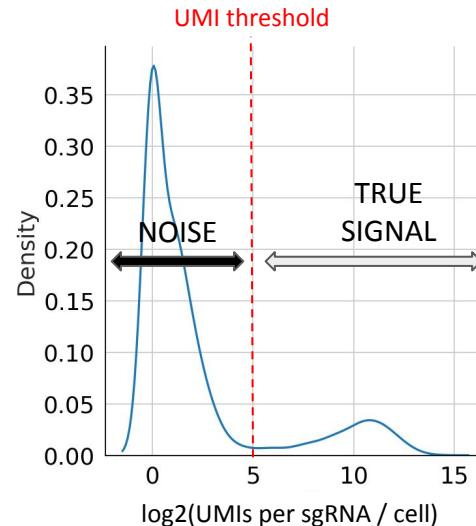
Algorithm:

- For each sgRNA
- Fit a 2-component Gaussian Mixture Model to the $\log(\text{UMI counts})$ across all cell-associated barcodes
- Cells that fall in the second mode called as having the sgRNA present

Which sgRNA induced an efficient target gene silencing ?

To assess sgRNA silencing efficiency, we need:

- a sufficient number of cells being infected by the sgRNA
- **being confident in the sgRNA assignment**



How sgRNA are assigned to cells ?

How to distinguish true signal from background protospacer UMIs ?

Cellranger sgRNA assignment call

UMI threshold is automatically computed by cellranger.

It is done per sgRNA, independent of all other sgRNAs.

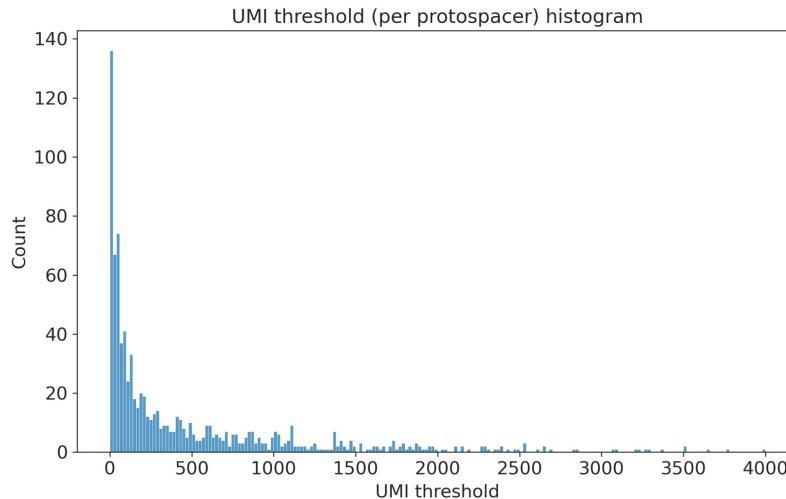
Algorithm:

- For each sgRNA
- Fit a 2-component Gaussian Mixture Model to the $\log(\text{UMI counts})$ across all cell-associated barcodes
- Cells that fall in the second mode called as having the sgRNA present

Which sgRNA induced an efficient target gene silencing ?

To assess sgRNA silencing efficiency, we need:

- a sufficient number of cells being infected by the sgRNA
- **being confident in the sgRNA assignment**



UMI threshold (one for each sgRNA) are very heterogeneous.

Very low UMI thresholds are suspicious.

How this automated assignment work with a very low number of cells ?

How sgRNA are assigned to cells ?

How to distinguish true signal from background protospacer UMIs ?

Cellranger sgRNA assignment call

UMI threshold is automatically computed by cellranger.

It is done per sgRNA, independent of all other sgRNAs.

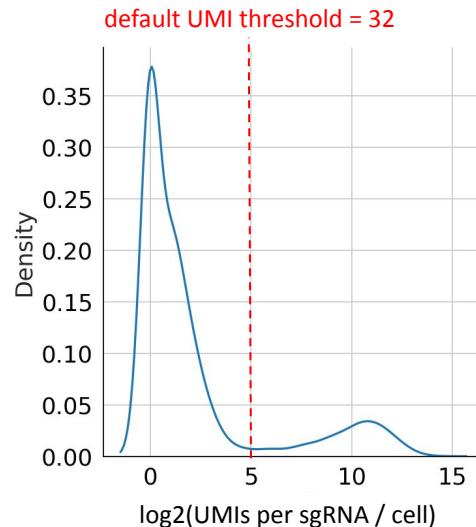
Algorithm:

- For each sgRNA
- Fit a 2-component Gaussian Mixture Model to the log(UMI counts) across all cell-associated barcodes
- Cells that fall in the second mode called as having the sgRNA present

Which sgRNA induced an efficient target gene silencing ?

To assess sgRNA silencing efficiency, we need:

- a sufficient number of cells being infected by the sgRNA
- **being confident in the sgRNA assignment**



How sgRNA are assigned to cells ?

How to distinguish true signal from background protospacer UMIs ?

Our sgRNA assignment call

Done per sgRNA, independent of all other sgRNAs.

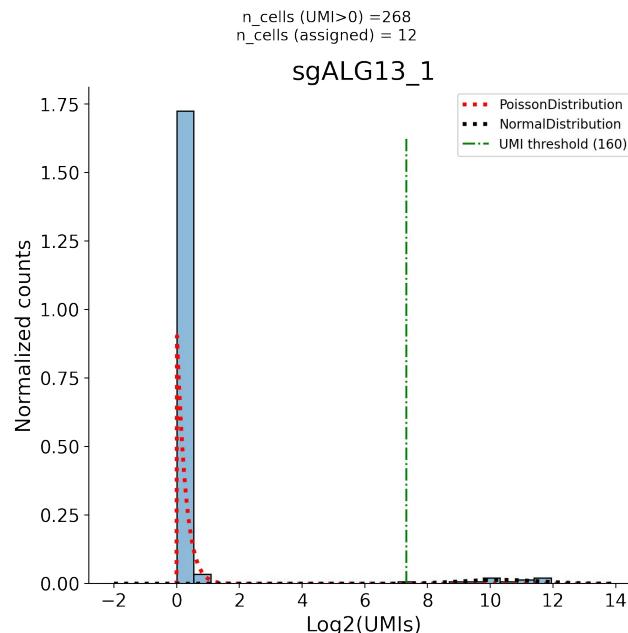
Algorithm:

- For each sgRNA
- Fit a 2-component **Poisson-Gaussian** Mixture Model to the $\log(\text{UMI counts})$ across all cell-associated barcodes
- Cells that fall in the second mode called as having the sgRNA present, if the **fitting converged and UMI threshold > 32**
- otherwise, apply default UMI threshold = 32

Which sgRNA induced an efficient target gene silencing ?

To assess sgRNA silencing efficiency, we need:

- a sufficient number of cells being infected by the sgRNA
- **being confident in the sgRNA assignment**



How sgRNA are assigned to cells ?

How to distinguish true signal from background protospacer UMIs ?

Our sgRNA assignment call

Done per sgRNA, independent of all other sgRNAs.

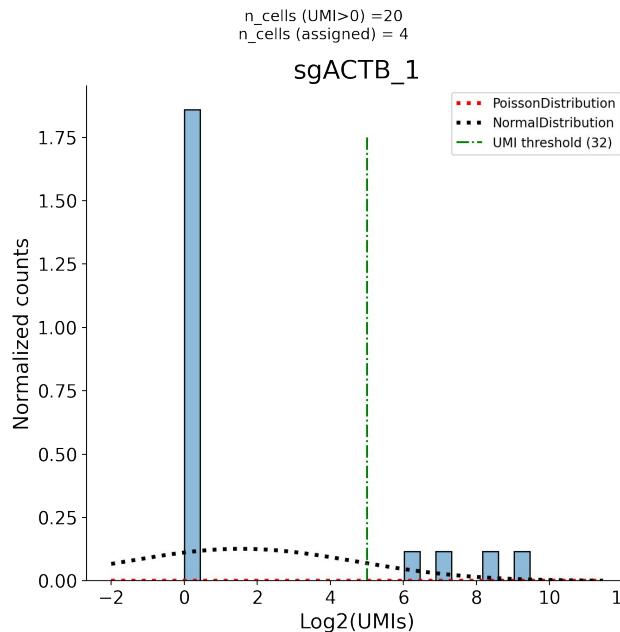
Algorithm:

- For each sgRNA
- Fit a 2-component **Poisson-Gaussian Mixture Model** to the log(UMI counts) across all cell-associated barcodes
- Cells that fall in the second mode called as having the sgRNA present, if the **fitting converged and UMI threshold > 32**
- otherwise, apply default UMI threshold = 32

Which sgRNA induced an efficient target gene silencing ?

To assess sgRNA silencing efficiency, we need:

- a sufficient number of cells being infected by the sgRNA
- **being confident in the sgRNA assignment**



How sgRNA are assigned to cells ?

How to distinguish true signal from background protospacer UMIs ?

Our sgRNA assignment call

Done per sgRNA, independent of all other sgRNAs.

Algorithm:

- For each sgRNA
- Fit a 2-component **Poisson-Gaussian** Mixture Model to the log(UMI counts) across all cell-associated barcodes
- Cells that fall in the second mode called as having the sgRNA present, if the **fitting converged and UMI threshold > 32**
- **otherwise, apply default UMI threshold = 32**

Which sgRNA induced an efficient target gene silencing ?

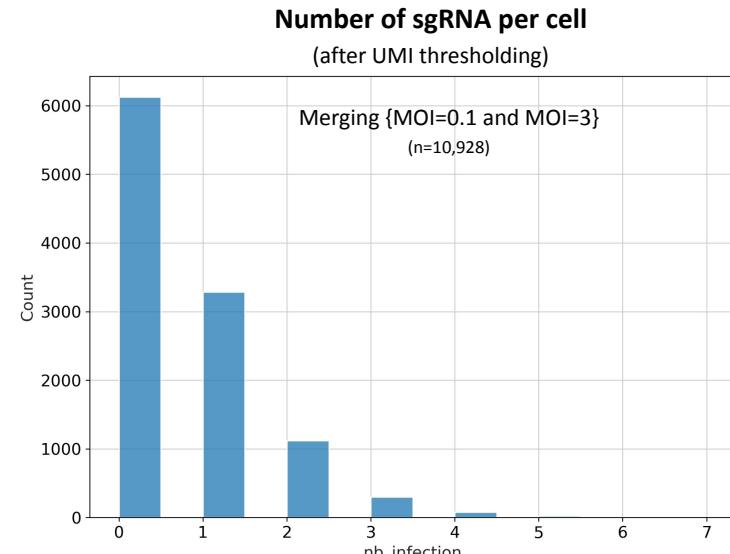
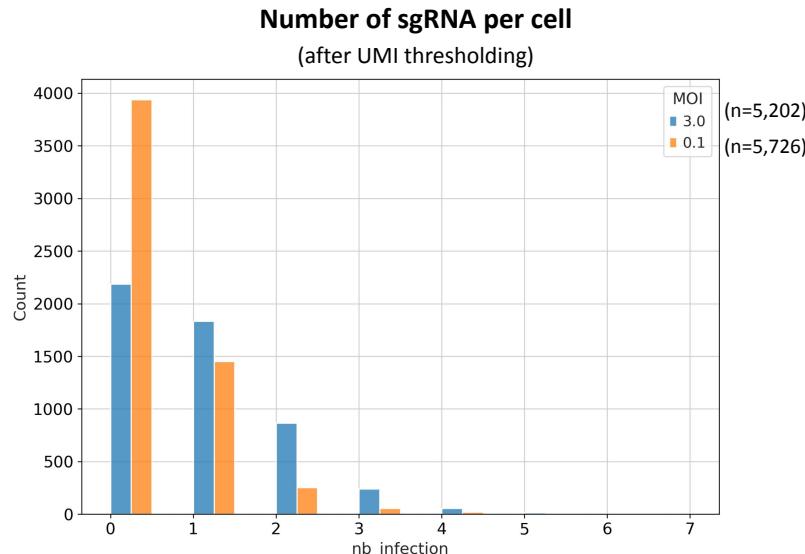
To assess sgRNA silencing efficiency, we need:

- a sufficient number of cells being infected by the sgRNA
- **being confident in the sgRNA assignment**

How sgRNA are assigned to cells ?

How to distinguish true signal from background protospacer UMIs ?

Our sgRNA assignment call



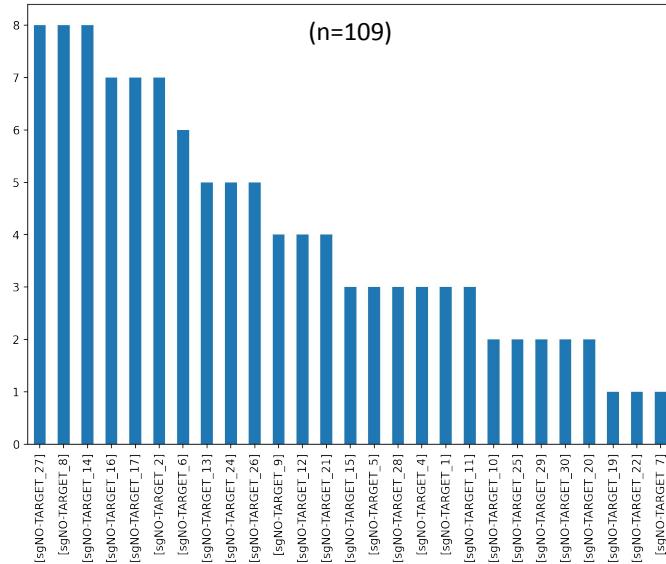
Which sgRNA induced an efficient target gene silencing ?

To assess sgRNA silencing efficiency, we need:

- a sufficient number of cells being infected by the sgRNA
- being confident in the sgRNA assignment
- **a control population to compare the target gene expression**

Untargeted cells = 6,123 (after UMI sgRNA threshold)

Numbers of cells infected by a Control (and only) sgRNA

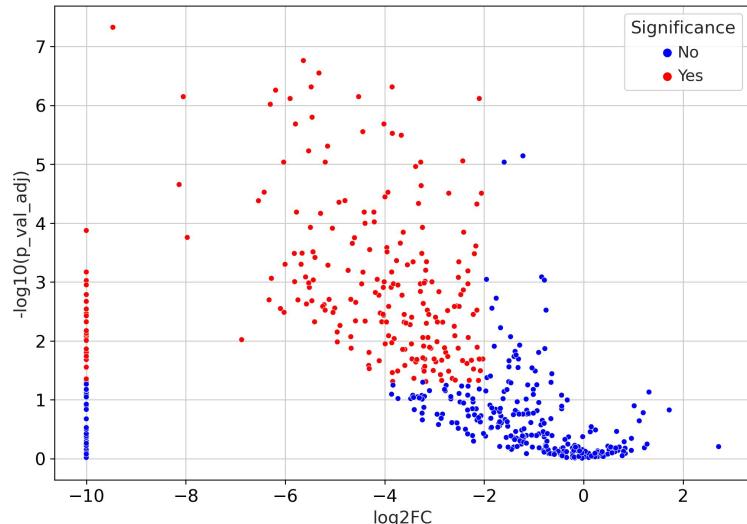


sgRNA silencing efficiency assessment

sgRNA is considered as efficiently downregulating target gene if:

- at least 5 cells are infected by the sgRNA
- & $\log_{2}\text{FC} < -2$ (compared to Control cells)
- & $p_{\text{val}} < 0.05$ (Wilcoxon test, BH correction)

sgRNAs inducing a significant downregulation of the target gene

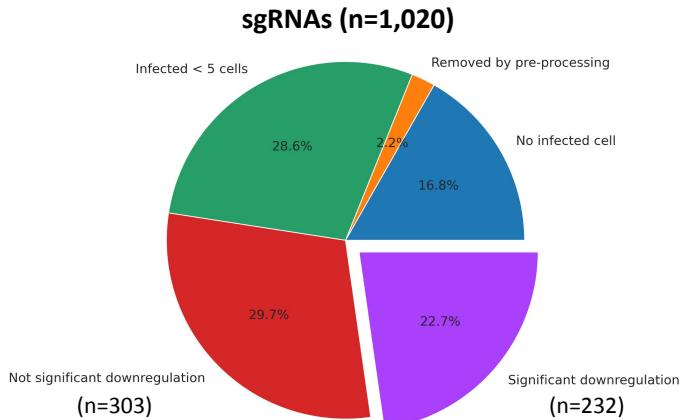
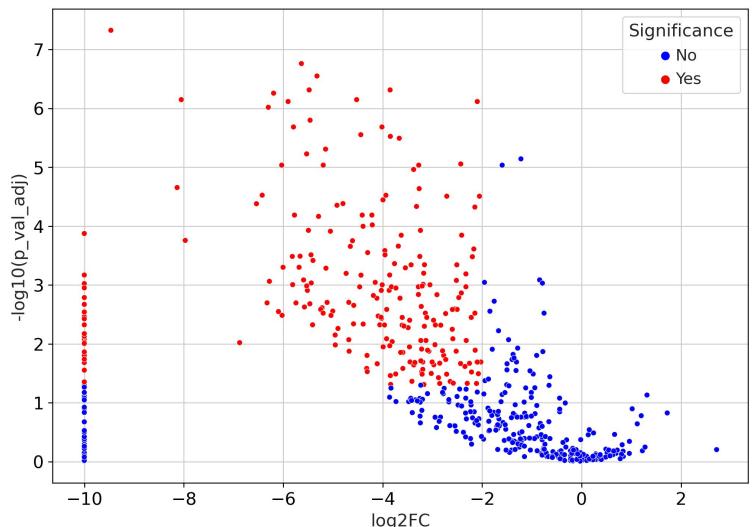


sgRNA silencing efficiency assessment

sgRNA is considered as efficiently downregulating target gene if:

- at least 5 cells are infected by the sgRNA
- & $\log_{2}FC < -2$ (compared to Control cells)
- & $p_{\text{val}} < 0.05$ (Wilcoxon test, BH correction)

sgRNAs inducing a significant downregulation of the target gene

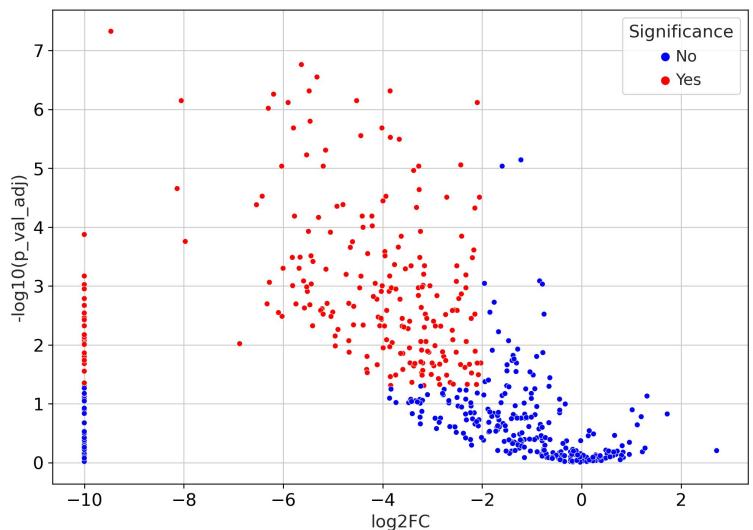


sgRNA silencing efficiency assessment

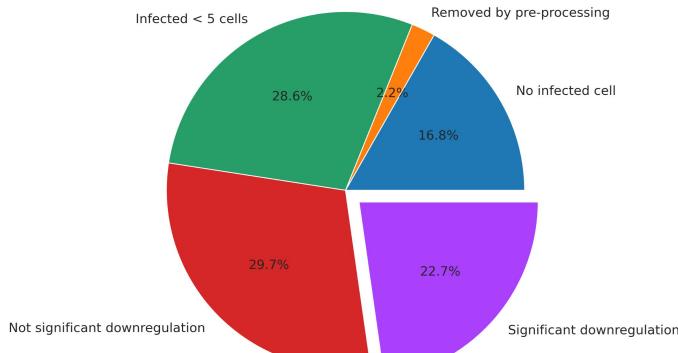
sgRNA is considered as efficiently downregulating target gene if:

- at least 5 cells are infected by the sgRNA
- & $\log_{2}FC < -2$ (compared to Control cells)
- & $p_{\text{val}} < 0.05$ (Wilcoxon test, BH correction)

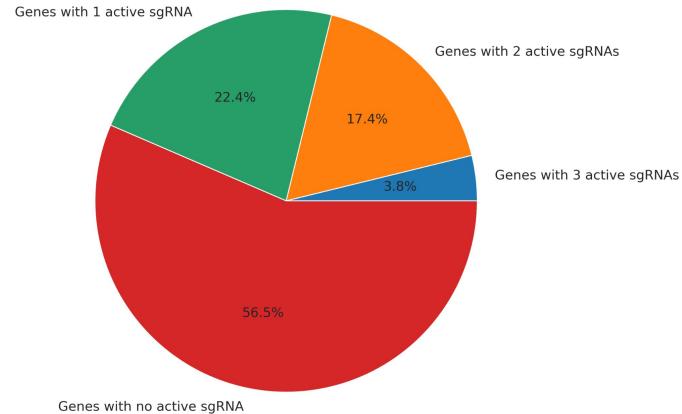
sgRNAs inducing a significant downregulation of the target gene



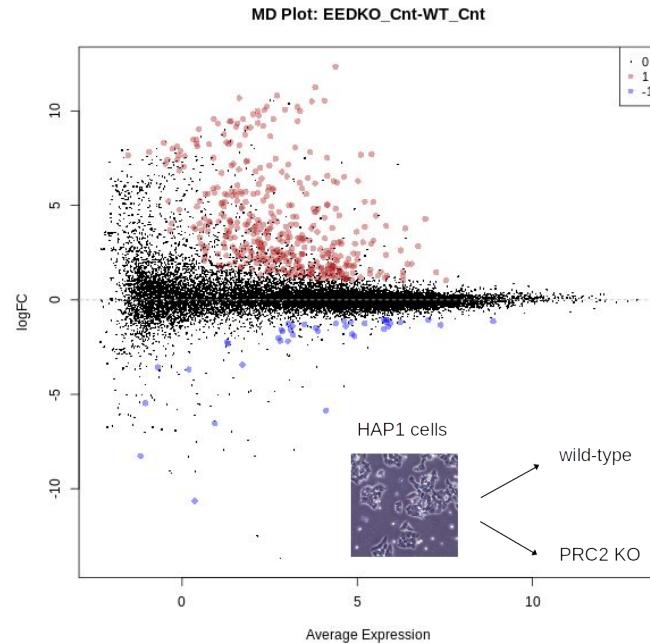
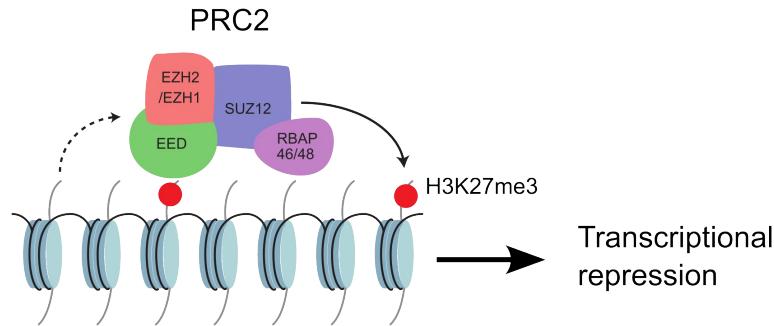
sgRNAs (n=1,020)



Target genes (n=340)

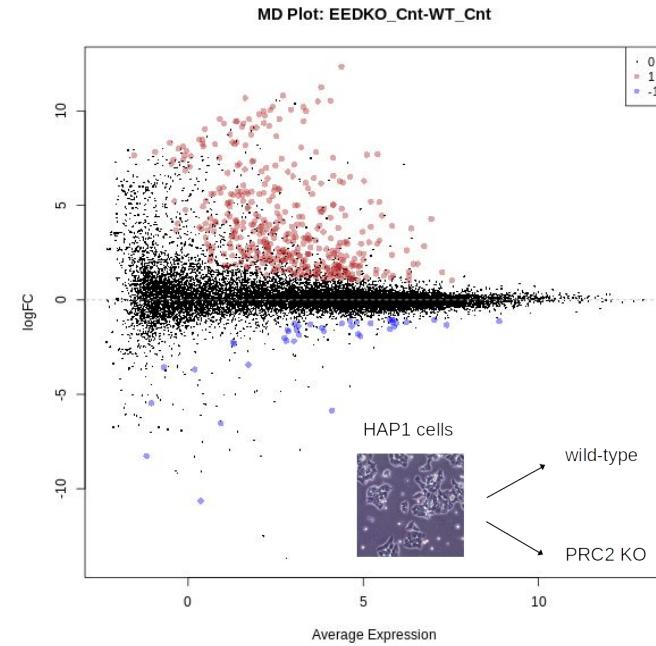
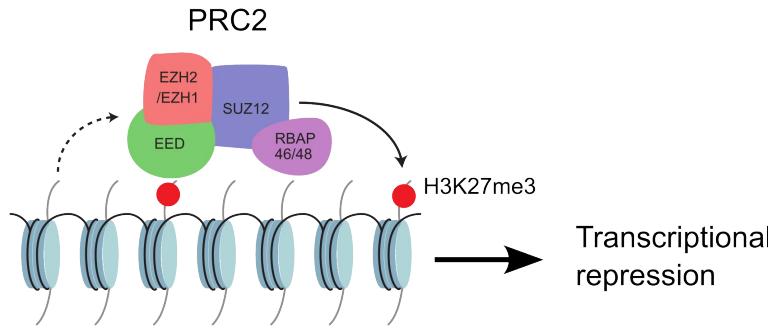


Quality control: Polycomb Repressive Complex 2 (PRC2)



bulk RNA-seq > identification of
316 upregulated genes in PRC2 KO cells

Quality control: Polycomb Repressive Complex 2 (PRC2)



bulk RNA-seq > identification of
316 upregulated genes in PRC2 KO cells

When looking at cells infected by an efficient sgRNA targeting a core member of the Polycomb PRC2 complex, do we observe upregulation of genes normally regulated by PRC2? (expected)

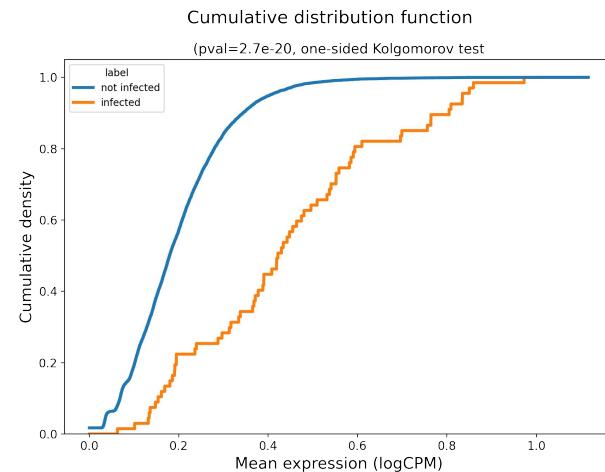
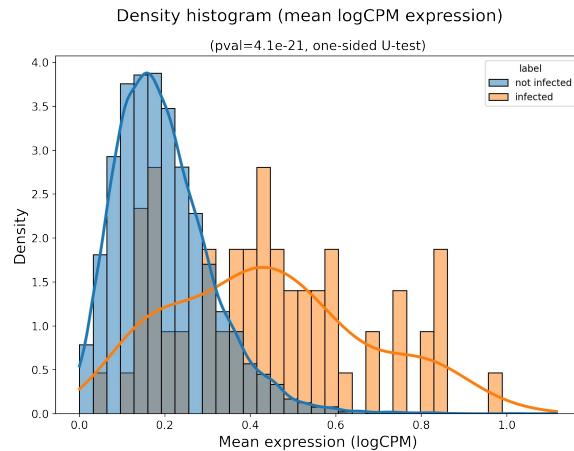
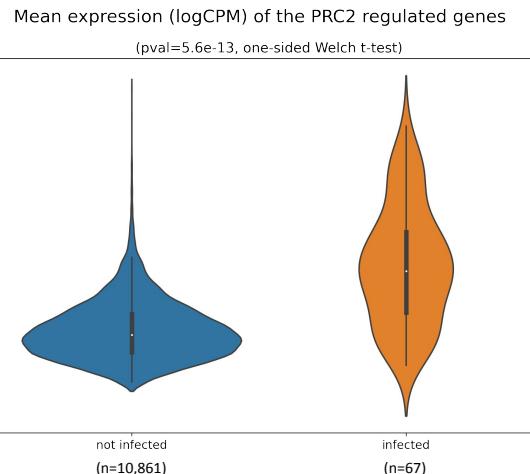
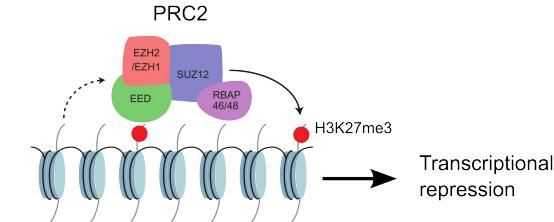
Quality control: Polycomb Repressive Complex 2 (PRC2)

n=67

(sgSUZ12_2, sgSUZ12_3, sgEZH2_1, sgEZH2_3)

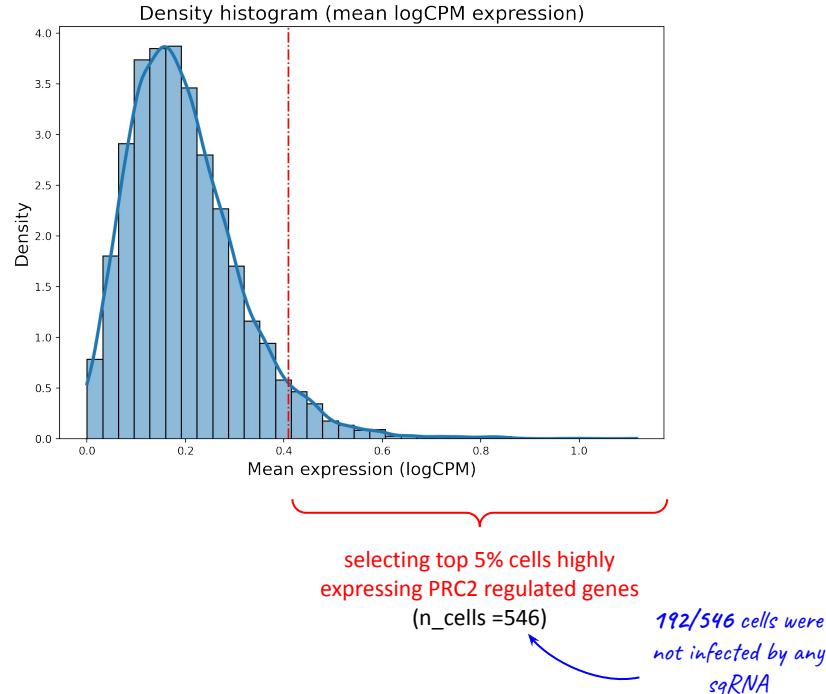
When looking at cells infected by an efficient sgRNA targeting a core member of the Polycomb PRC2 complex, do we observe upregulation of genes normally regulated by PRC2 ? (expected)

316 upregulated genes in PRC2 KO cells



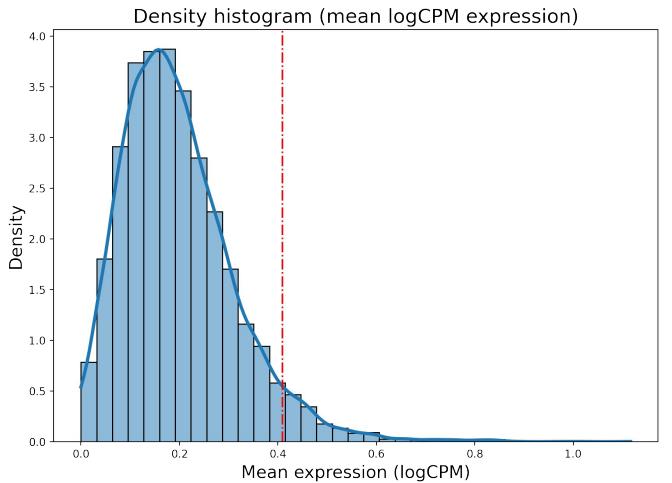
Quality control: Polycomb Repressive Complex 2 (PRC2)

When looking at cells that upregulated the most genes
regulated by PRC2, by which sgRNAs have they been infected ?

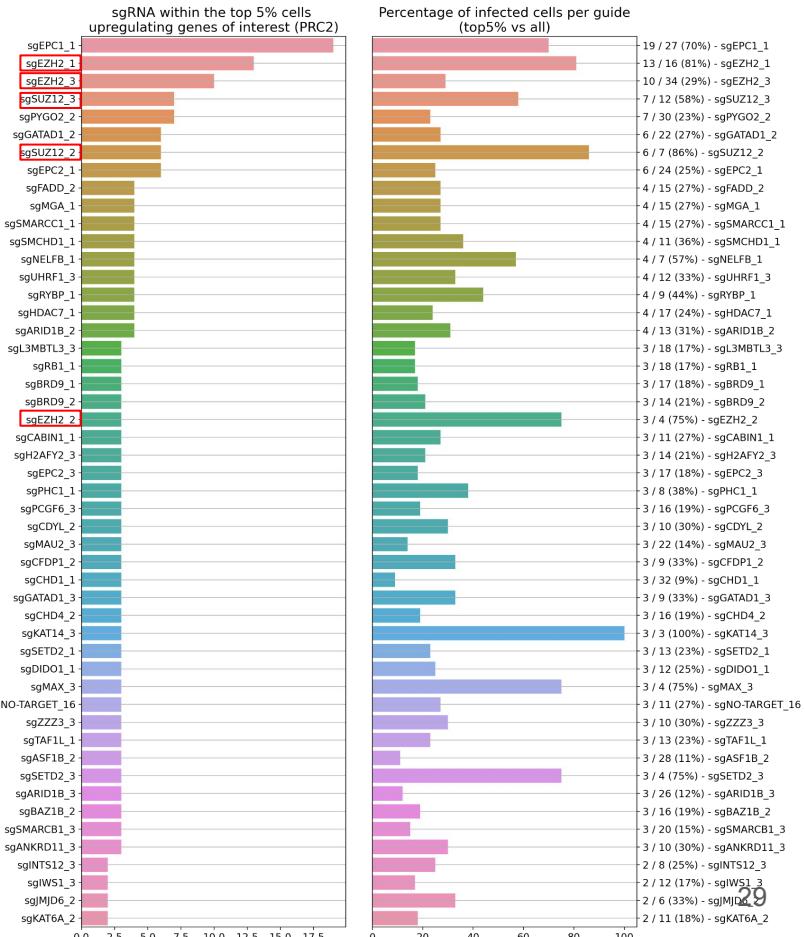


Quality control: Polycomb Repressive Complex 2 (PRC2)

When looking at cells that upregulated the most genes regulated by PRC2, by which sgRNAs have they been infected ?



192/546 cells were not infected by any sgRNA



Conclusion and next steps

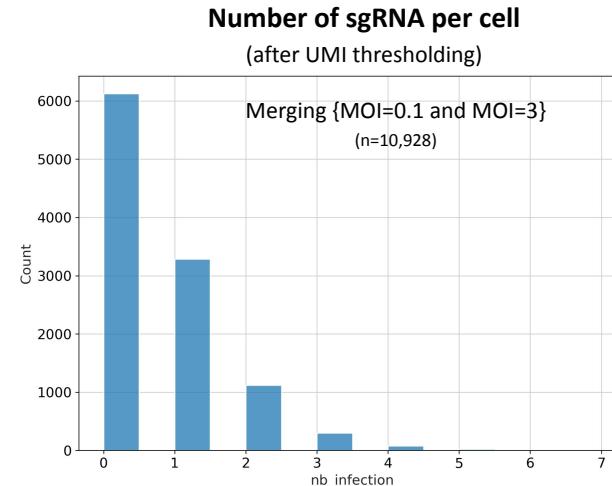
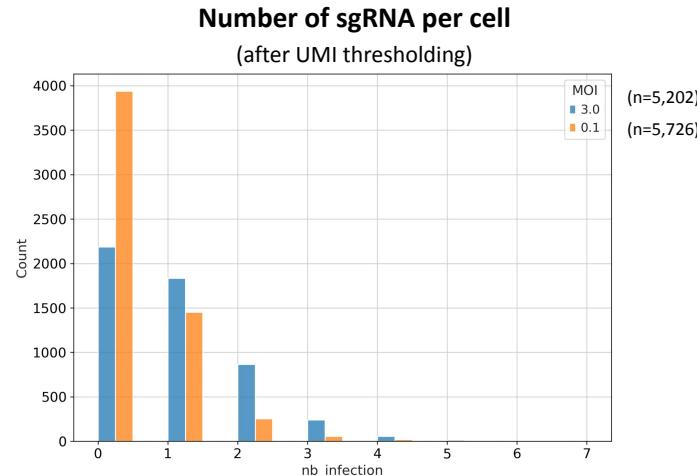
- We applied stringent UMI thresholding for sgRNA assignments
- Among sgRNAs infecting at least 5 cells, ~40% show a significant silencing efficacy

Conclusion and next steps

- We applied stringent UMI thresholding for sgRNA assignments
- Among sgRNAs infecting at least 5 cells, ~40% show a significant silencing efficacy
- Relaunch another preliminary experiment to extend the list of sgRNAs that infected a minimal number of cells ?
- Real functional experiment using the best identified sgRNAs (using 2 efficient sgRNAs per lentivirus ?), with MOI~0.1 to limit number of multiple infections per cell.

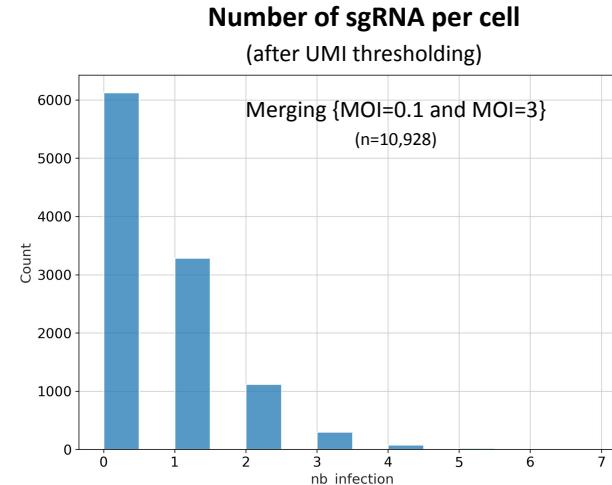
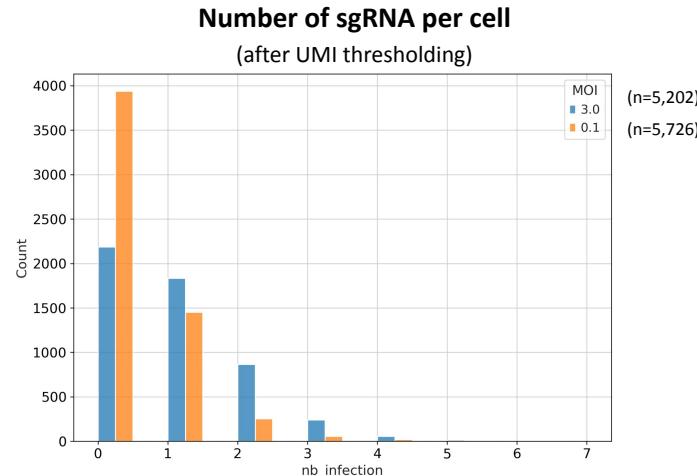
Conclusion and next steps

- We applied stringent UMI thresholding for sgRNA assignments
- Among sgRNAs infecting at least 5 cells, ~40% show a significant silencing efficacy
- Relaunch another preliminary experiment to extend the list of sgRNAs that infected a minimal number of cells ?
- Real functional experiment using the best identified sgRNAs (using 2 efficient sgRNAs per lentivirus ?), with MOI \sim 0.1 to limit number of multiple infections per cell.
- **A high fraction of potential CRISPR feature reads are not recognized during cellranger alignment**



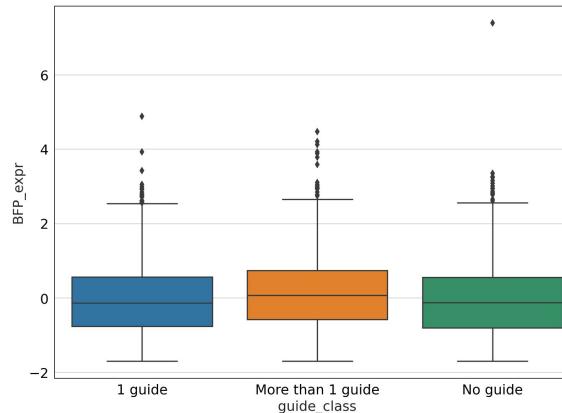
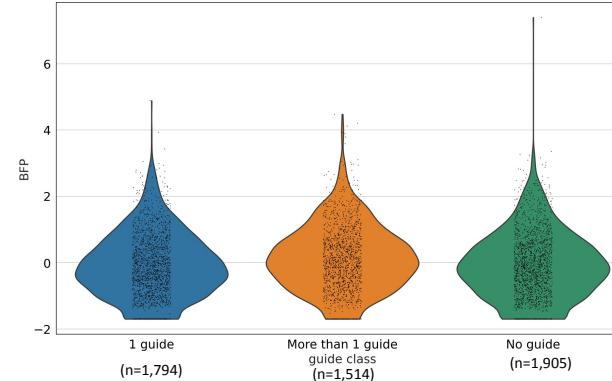
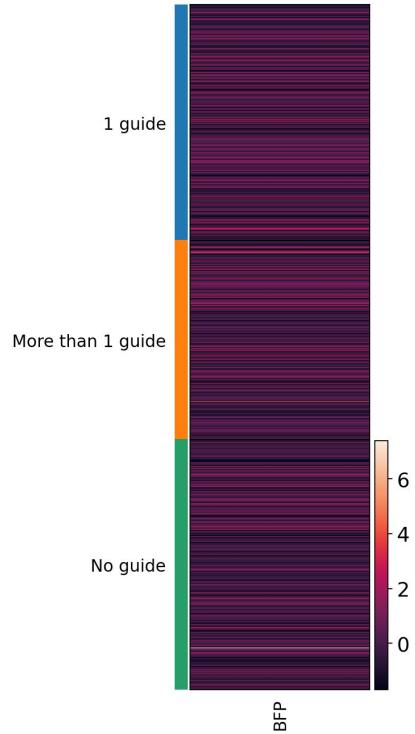
Conclusion and next steps

- We applied stringent UMI thresholding for sgRNA assignments
- Among sgRNAs infecting at least 5 cells, ~40% show a significant silencing efficacy
- Relaunch another preliminary experiment to extend the list of sgRNAs that infected a minimal number of cells ?
- Real functional experiment using the best identified sgRNAs (using 2 efficient sgRNAs per lentivirus ?), with MOI \sim 0.1 to limit number of multiple infections per cell.
- **A high fraction of potential CRISPR feature reads are not recognized during cellranger alignment**



Problems in FACS sorting ?

A high fraction of potential CRISPR feature reads are not recognized during cellranger alignment

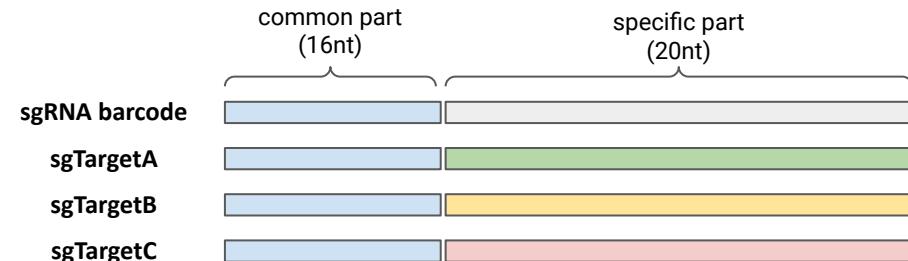


Problems in feature reads alignment ?

A high fraction of potential CRISPR feature reads are not recognized during cellranger alignment

CRISPR Application ②	
Fraction Reads with Putative Protospacer Sequence	37.1%
Fraction Guide Reads	25.8%
Fraction Guide Reads Usable	21.6%
Guide Reads Usable per Cell	6,673
Fraction Protospacer Not Recognized	30.4%
Guide Reads in Cells	84.4%
Cells with 1 or more protospacers detected	76.8%
Cells with 2 or more protospacers detected	51.2%
Median UMIs per Cell	279

from Cellranger output summary



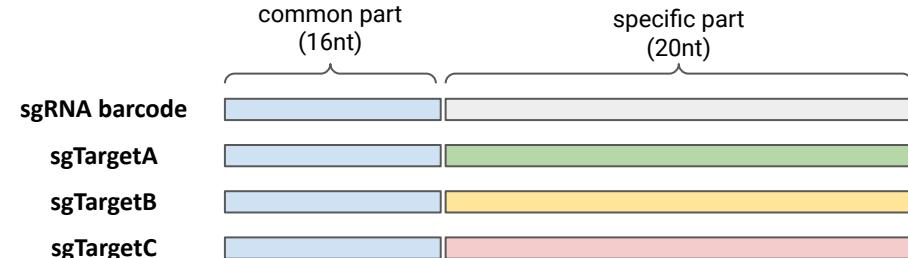
Problems in feature reads alignment ?

A high fraction of potential CRISPR feature reads are not recognized during cellranger alignment

CRISPR Application ②	
Fraction Reads with Putative Protospacer Sequence	37.1%
Fraction Guide Reads	25.8%
Fraction Guide Reads Usable	21.6%
Guide Reads Usable per Cell	6,673
Fraction Protospacer Not Recognized	30.4%
Guide Reads in Cells	84.4%
Cells with 1 or more protospacers detected	76.8%
Cells with 2 or more protospacers detected	51.2%
Median UMIs per Cell	279

from Cellranger output summary

“xf:i:24” (24=8+16)
are the reads that are
kept for sgRNA
assignment



Cellranger adds different metadata in the bam alignment file

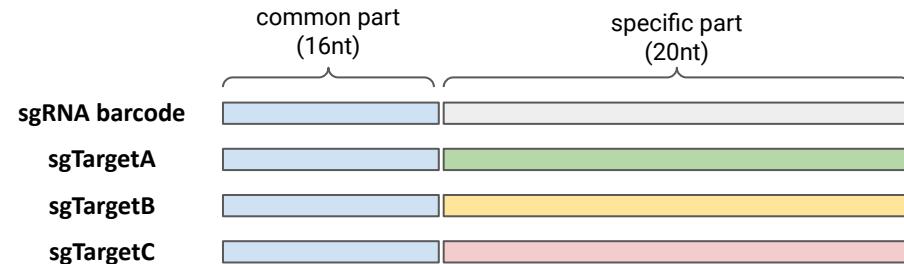
xf	i	Extra alignment flags. The bits of this tag are interpreted as follows: <ul style="list-style-type: none">• 1 - The read is confidently mapped to the transcriptome• 2 - This read's barcode, UMI, and feature combination was discarded in favor of a different feature with higher read support• 4 - This read pair maps to a discordant pair of genes, and is <i>not</i> treated as a UMI count• 8 - This read is representative for a molecule and is treated as a UMI count• 16 - This read maps to exactly one feature, and is identical to bit 1 for transcriptomic reads. Notably, this bit is set for a Feature Barcode read, while bit 1 is not• 32 - This read was removed by targeted UMI filtering.
----	---	---

<https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/output/bam>

Typos in the reference feature file ?

CRISPR Application ②	
Fraction Reads with Putative Protospacer Sequence	37.1%
Fraction Guide Reads	25.8%
Fraction Guide Reads Usable	21.6%
Guide Reads Usable per Cell	6,673
Fraction Protospacer Not Recognized	30.4%
Guide Reads in Cells	84.4%
Cells with 1 or more protospacers detected	76.8%
Cells with 2 or more protospacers detected	51.2%
Median UMIs per Cell	279

from Cellranger output summary



116 sgRNAs have not been found by cellranger in MOI=0.1 nor MOI=3