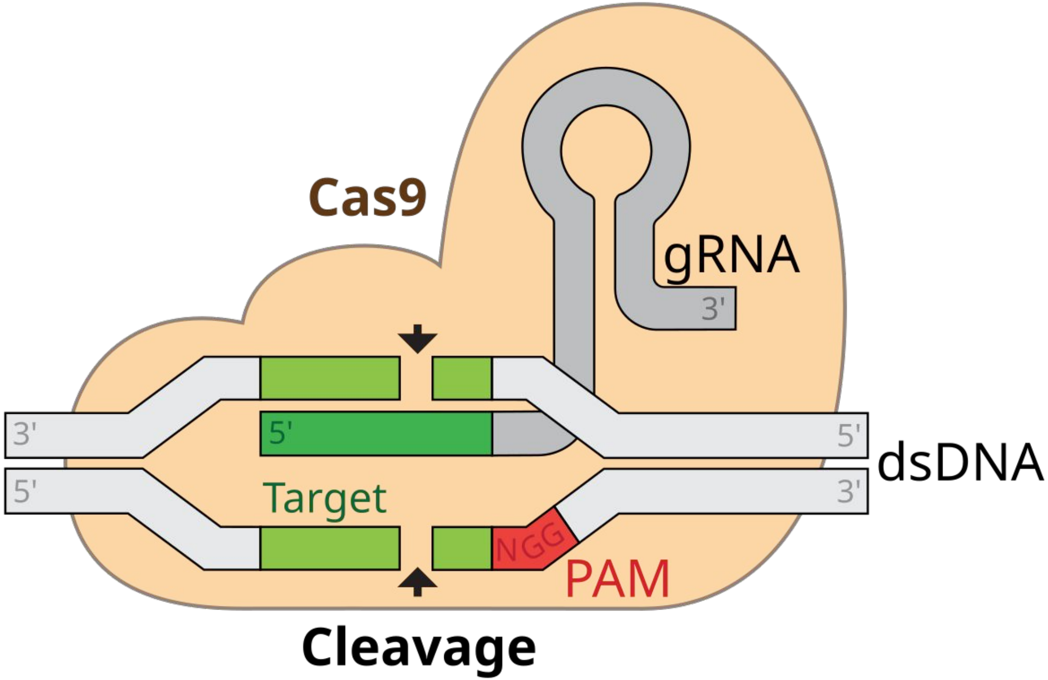
Capturing ultra-rare CRISPR-Cas9 off-target editing events in single cells

Fabrizio Malatesta
Graham McVicker, Brad Balderson
McVicker Lab, Salk Institute For Biological Sciences

UC LEADS
UCSD SRC 2025
8/13/2025

What is CRISPR-Cas9?

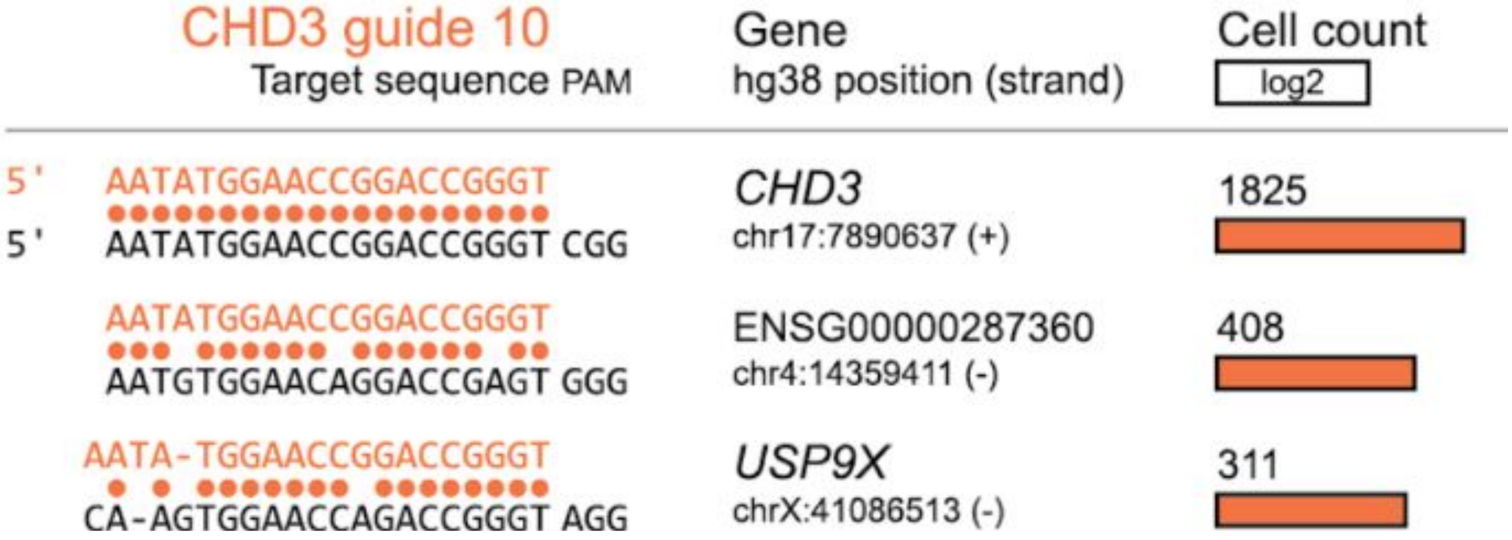
- Genome editing tool
- **Cas9** cuts DNA at specific sites
- A guide RNA (**gRNA**) leads Cas9 to the target
- Cas9 makes a double-strand break (**DSB**) near a **PAM (NGG or NAG)**
- DSBs repaired by imperfect repair mechanisms, which can introduce **indels**



Marius Walter, CC BY-SA 4.0 via Wikimedia Commons

CRISPR Often Cuts DNA at Unintended Sites

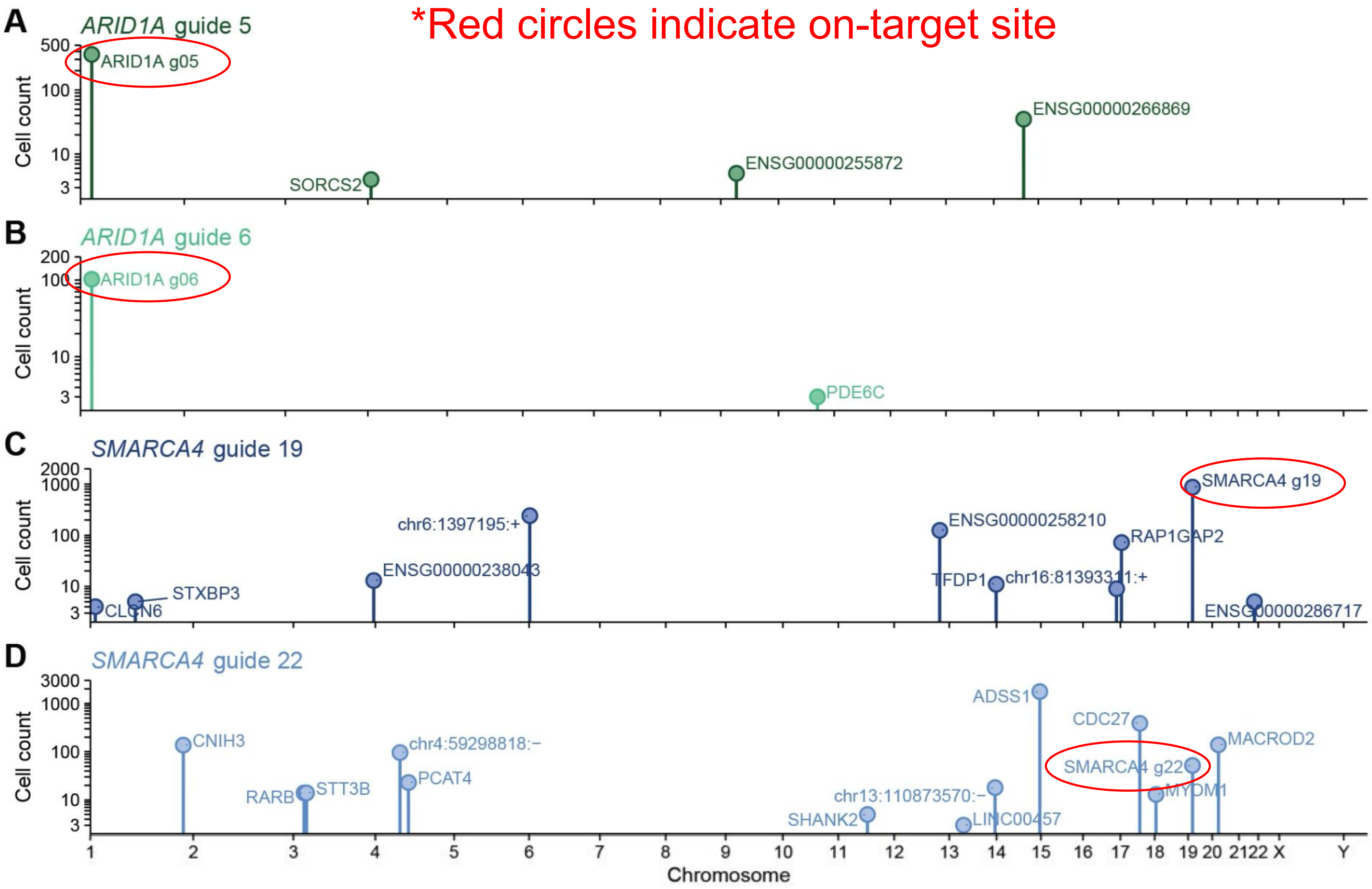
- CRISPR can cut unintended sites with similar sequences to the guide RNA



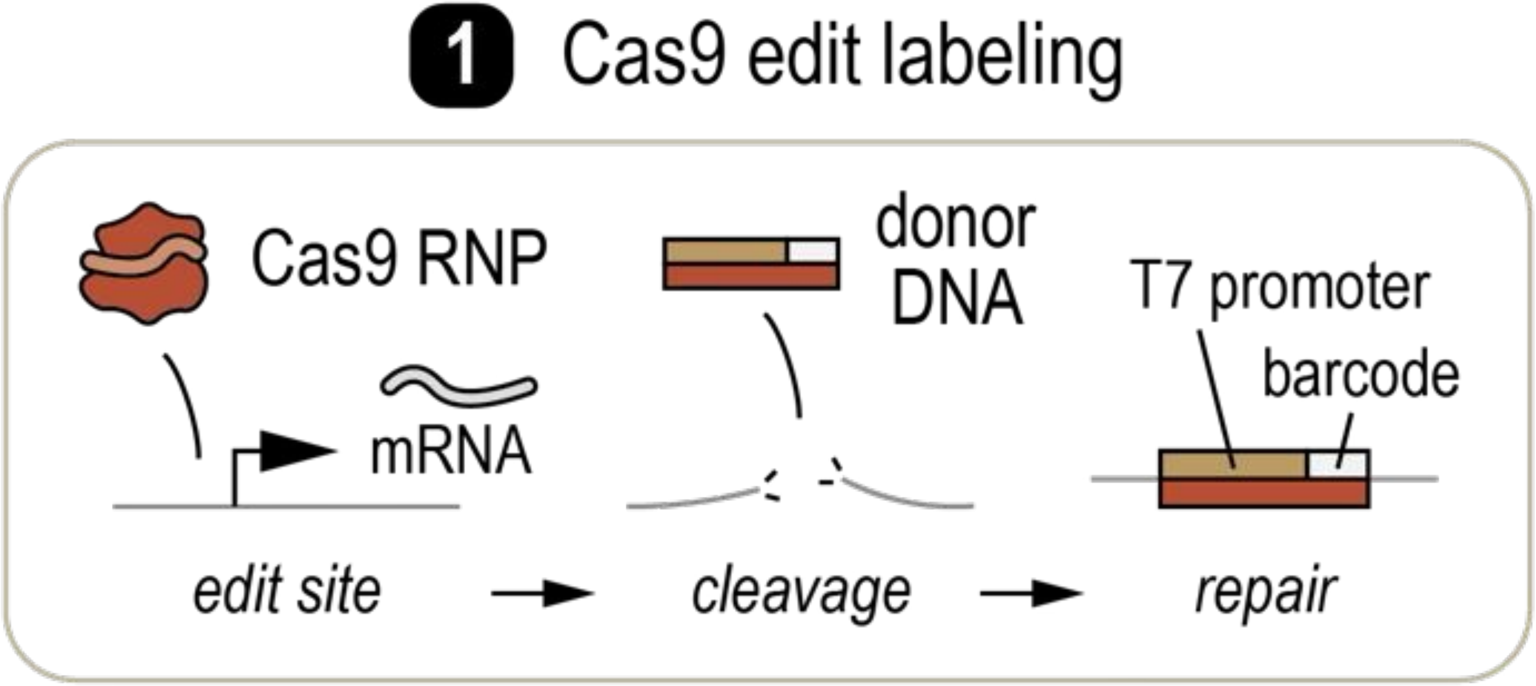
CHD3 guide 10 Target sequence PAM	Gene hg38 position (strand)	Cell count log2
5' AATATGGAACCGGACCGGGT ●●●●●●●●●●●●●●●●●●●● 5' AATATGGAACCGGACCGGGT CGG	*CHD3* chr17:7890637 (+)	1825
AATATGGAACCGGACCGGGT ●●● ●●●●●● ●●●●●● ●● AATGTGGAACAGGACCGAGT GGG	ENSG00000287360 chr4:14359411 (-)	408
AATA-TGGAACCGGACCGGGT ● ● ●●●●●●● ●●●●●●●●● CA-AGTGGAACCAGACCGGGT AGG	*USP9X* chrX:41086513 (-)	311

- Current tools are not effective at detecting/predicting off target edits

Enter Superb-seq

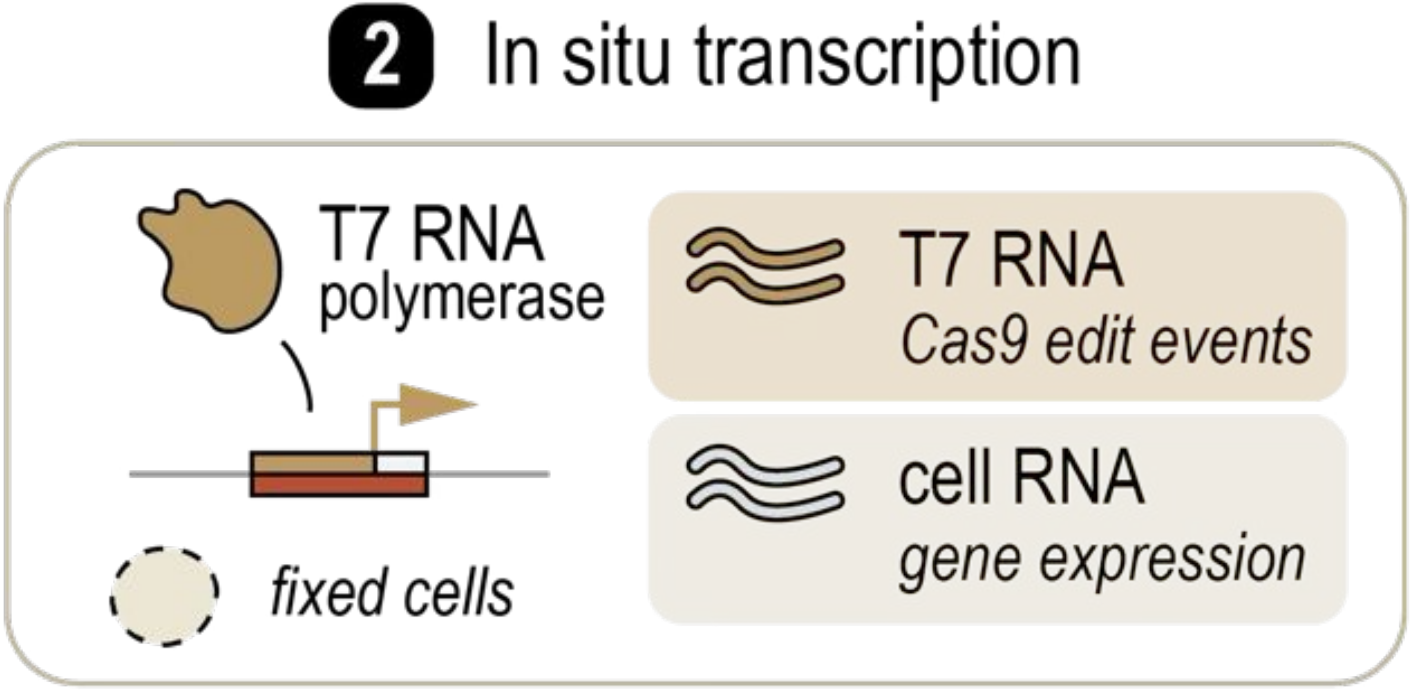

*Red circles indicate on-target site

Superb-seq Enables Single-Cell Detection of CRISPR Edits



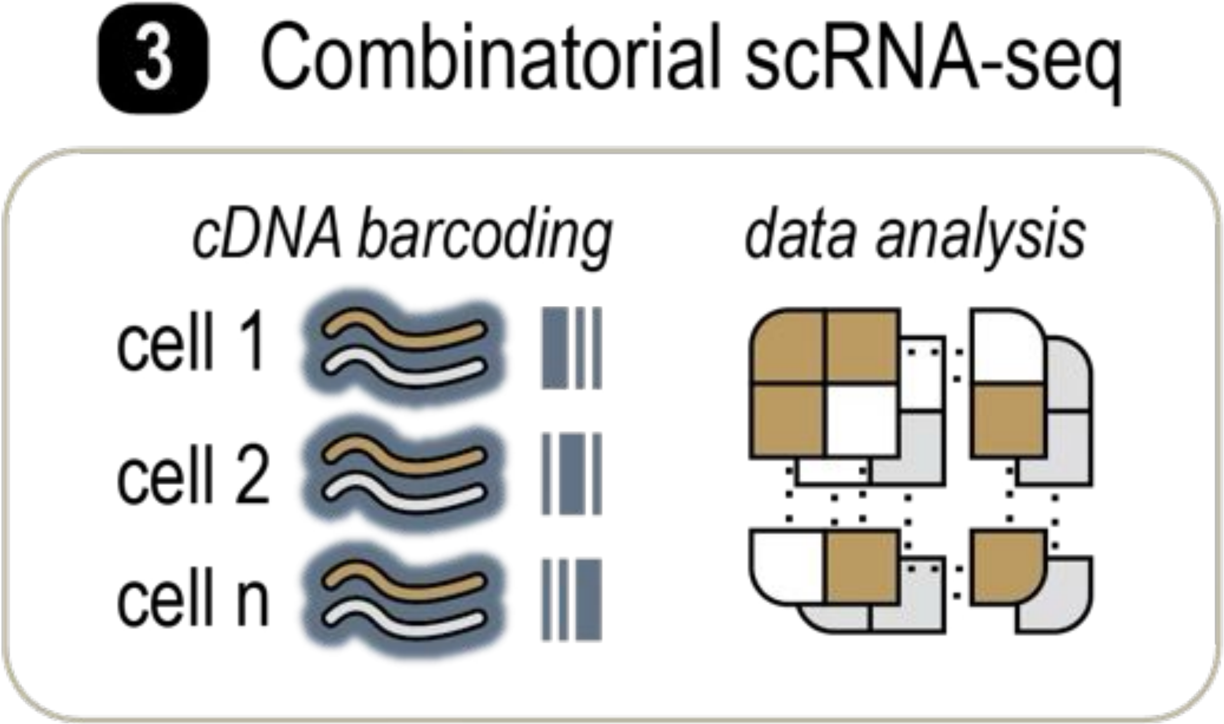
Homology-free labeling of on/off-targets

Superb-seq Enables Single-Cell Detection of CRISPR Edits



Single-cell genome amplification

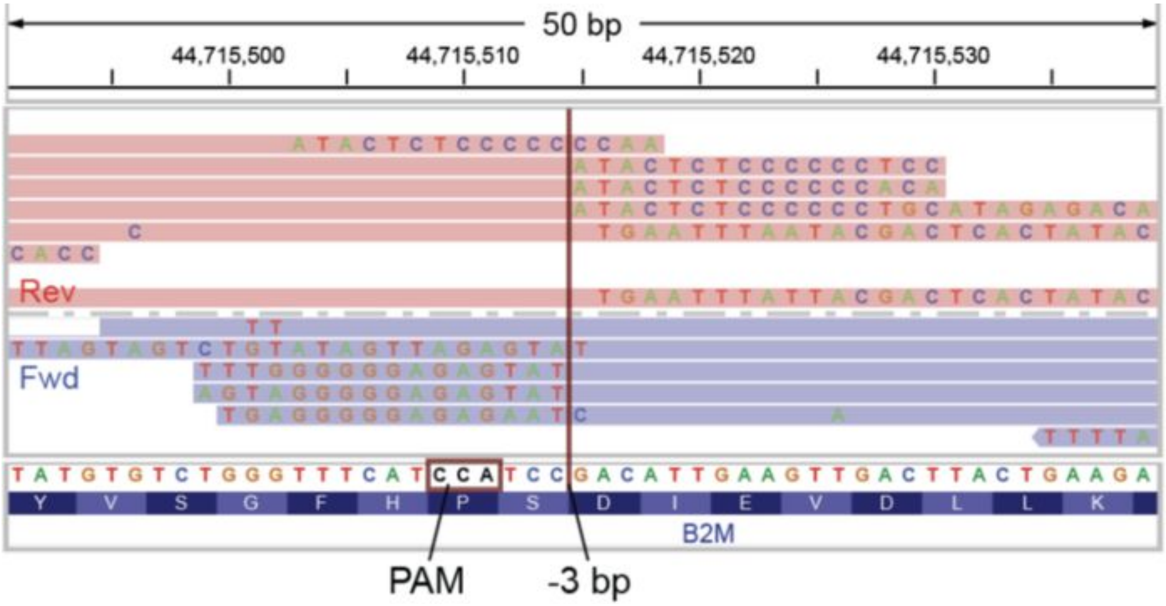
Superb-seq Enables Single-Cell Detection of CRISPR Edits



Off-the-shelf kits, custom software

Sheriff Identifies Candidate CRISPR Edits from Superb-Seq Data

- Software that identifies CRISPR edits by locating the T7 Barcode (cut site marker)
- Uses a set of filters to remove false positives
- Requires edits to appear in ≥3 cells + bidirectional support
- Downside: Misses ultra-rare (1–2 cell) edits



Project Goals and Hypothesis

Goal:

- Detect ultra-rare CRISPR editing events from Superb-seq data

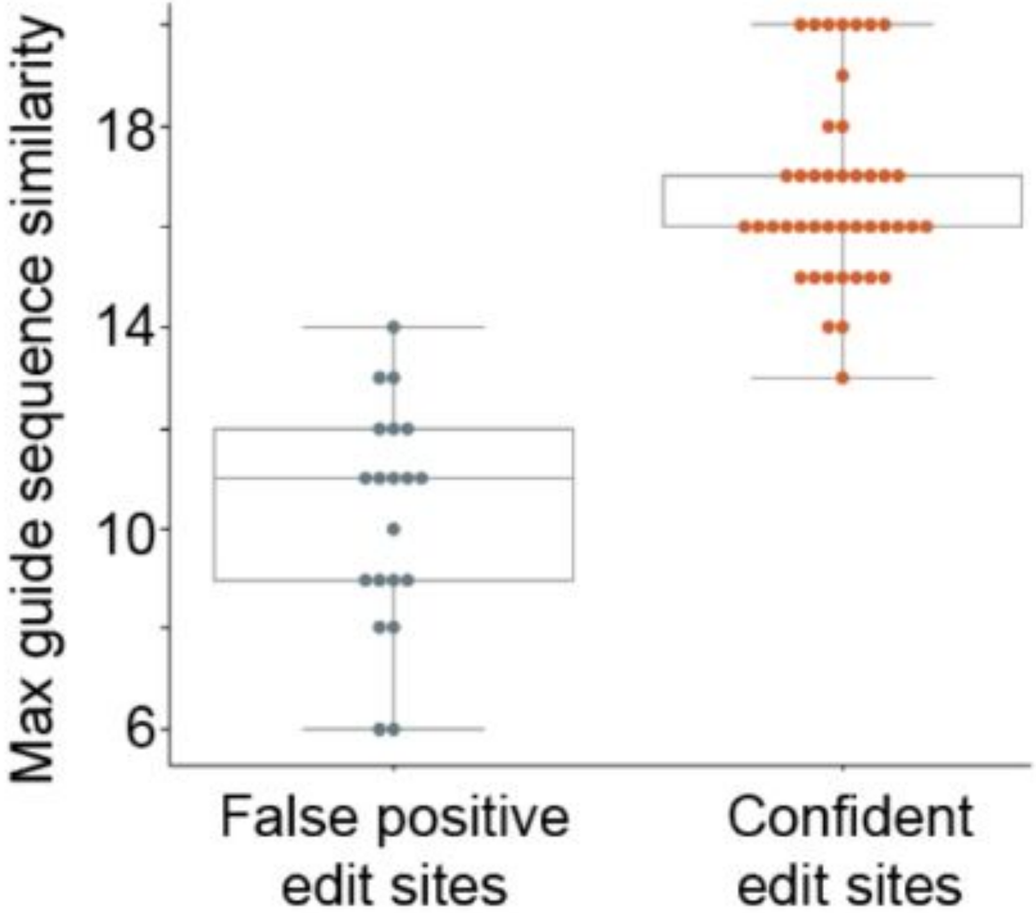
Why? To gain deeper insights into CRISPR editing and the interpretability of single-cell perturbation screens

Hypothesis:

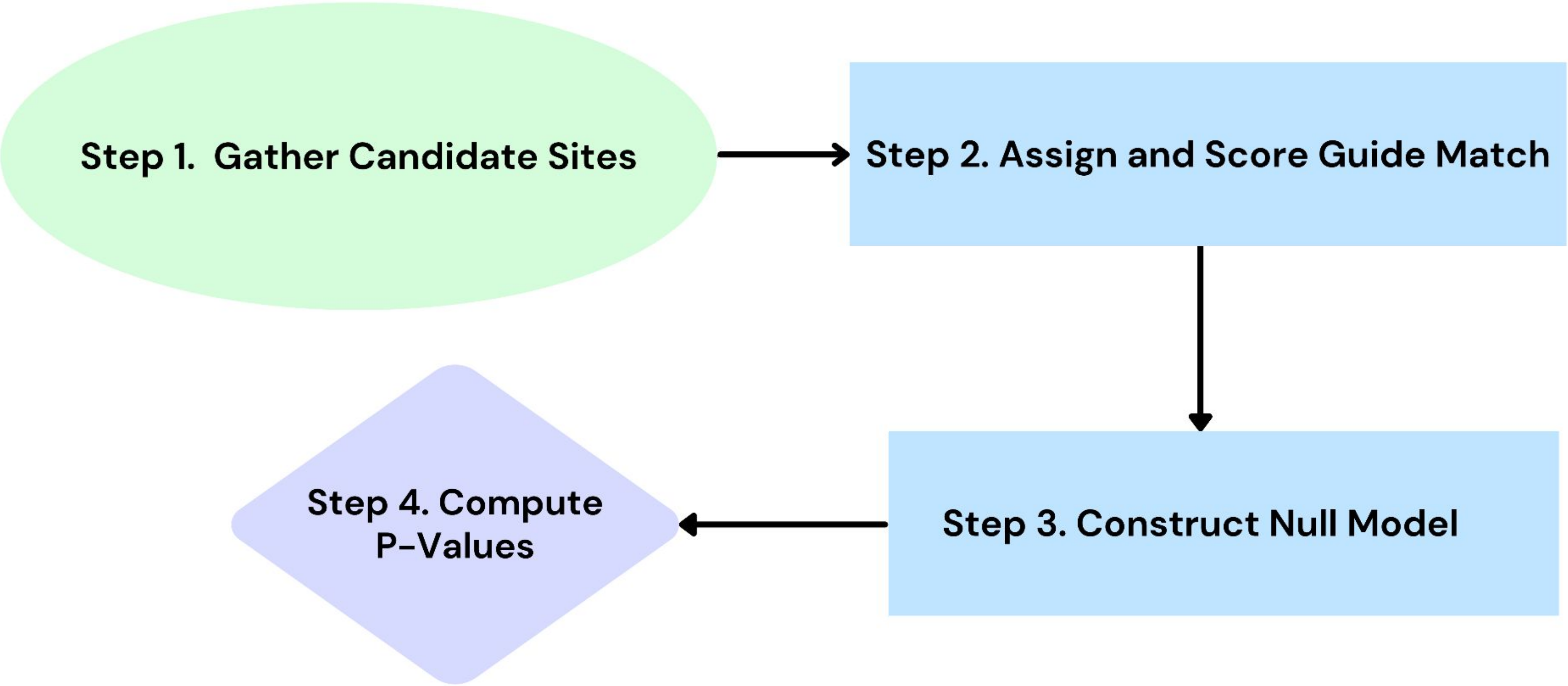
- I hypothesize that some of these ultra-rare candidate edits represent real CRISPR-Cas9 off-target events.

High False Positive Rate Even With Standard Filtering

- With standard filters set, 20/63 calls were still false positives
- Loosening the filters will only produce more



Methods Overview



Step 1. Gather Candidate Sites

Step 2. Assign and Score Guide Match

Step 3. Construct Null Model

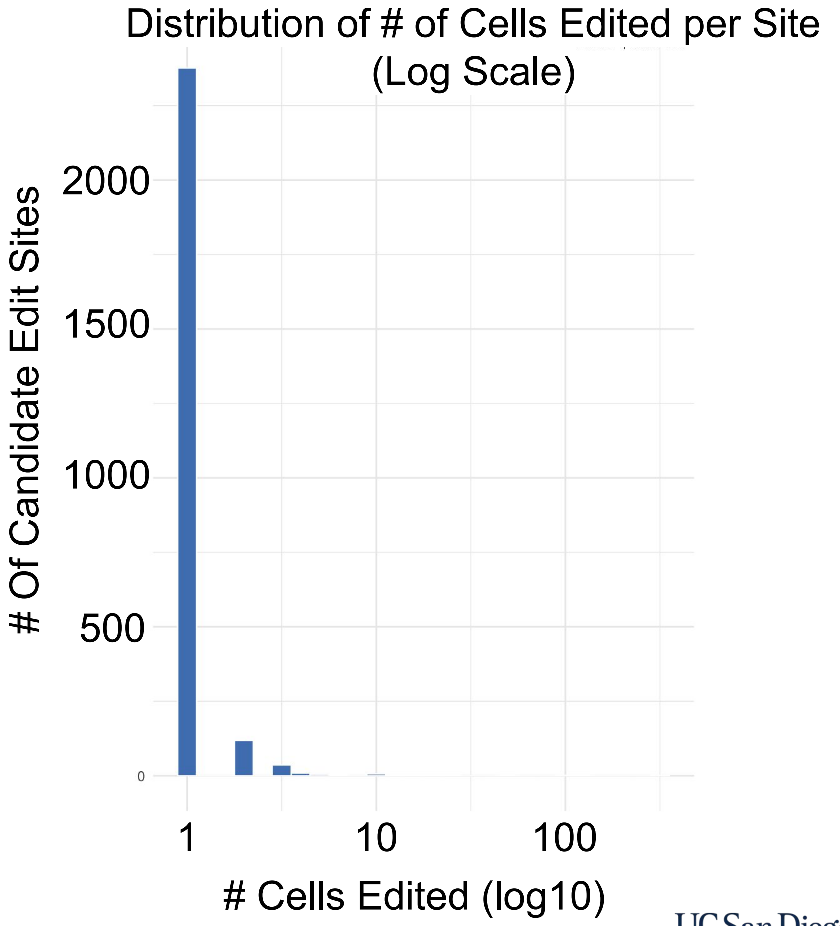
Step 4. Compute P-Values

Note: Only analyzed data from chromosome 1 to allow reasonable runtime

Relaxed Sheriff Reveals Thousands of Rare Candidate Edits

Sheriff Setting	**# Edits Detected**
Strict (default)	6
Relaxed (mine)	2562

- Problem: How do we differentiate the quality of all these new edit sites?



Identifying Likely Off-Targets via Guide Homology Scoring

- **Candidate Site to Guide Homology Scoring**

- **Method**:
  - Extract **10bp window** around candidate site
  - identify all **valid PAM (NGG or NAG) sequences**
  - score guides against the **adjacent 20 bp** with **global sequence alignment**:
    - base match = +1
    - mismatch = -1
    - gap open = -2
    - gap extend = -0.5

Guides Considered:

- ARID1A g05
- ARID1A g06
- SMARCA4 g19
- SMARCA4 g22

SMARCA4 guide 22 Target sequence PAM	Gene hg38 position (strand)
ACCGAGGGTGGAGGTAGCGT ACCGAGGGTGGAGGTAGCGT TGG	*SMARCA4* chr19:10986484 (-)
ACCGAGGGTGGAGGTAGCGT TGGGAGGGAGGAGGTAGCGT CGG	*ADSS1* chr14:104729483 (+

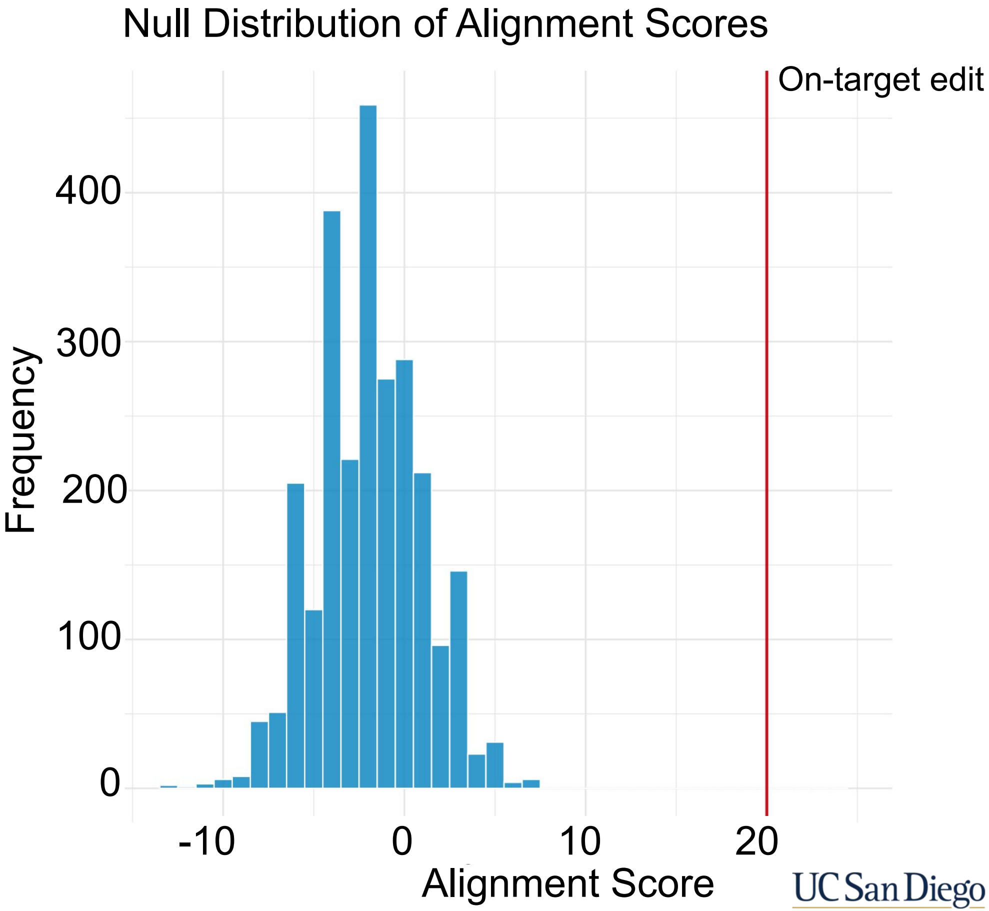
Building a Null Model to Evaluate Off-Target Evidence

Problem:

- Many guide-edit matches could happen by random sequence similarity.

Solution:

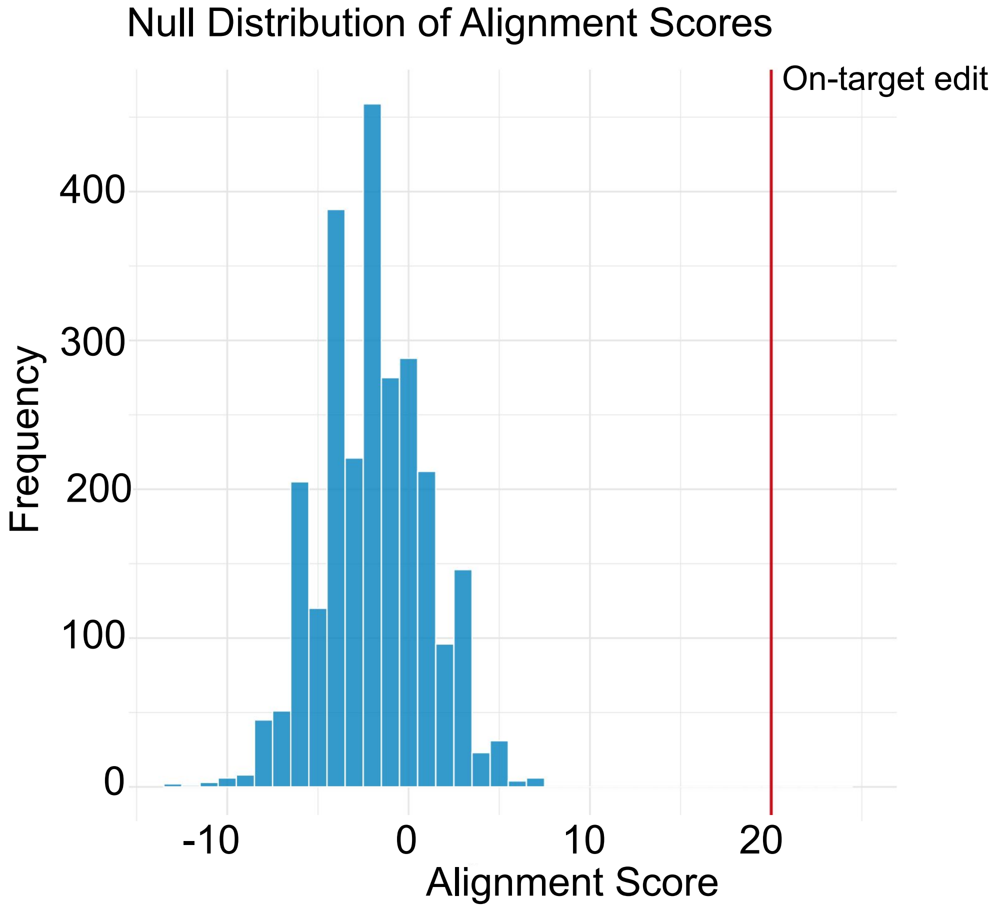
- Build null model
  - Scramble guide sequences
  - Rerun guide homology scoring



UC San Diego

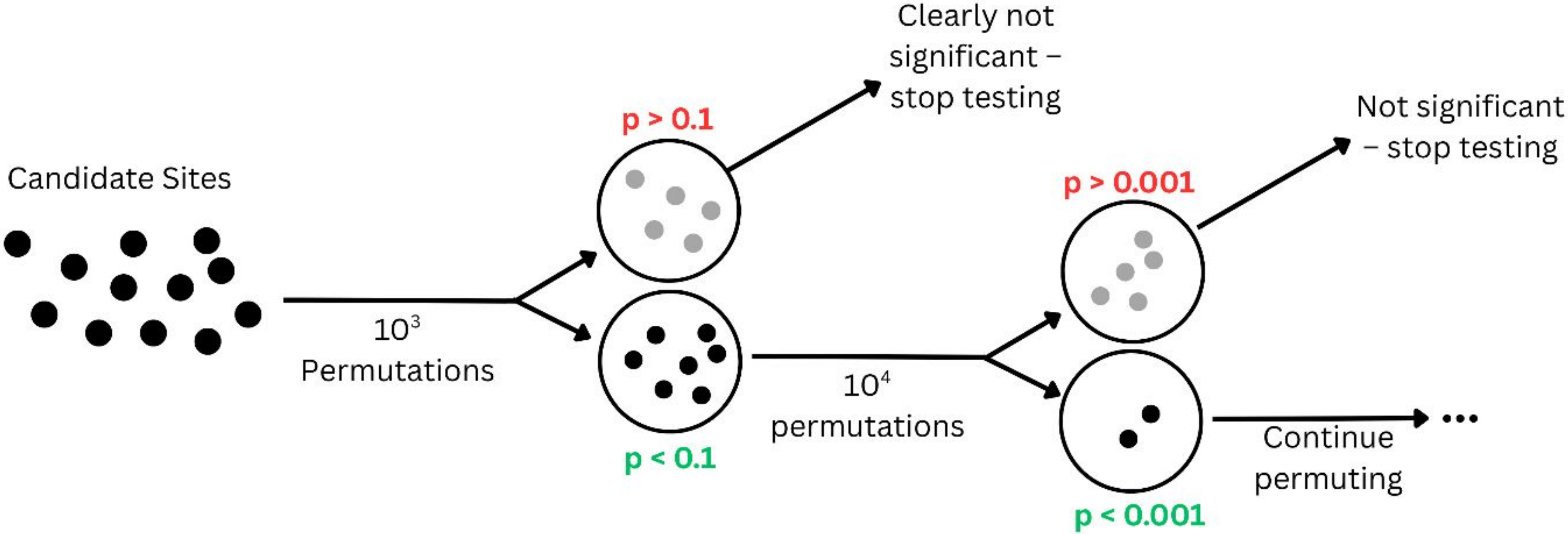
Empirical P-Value

Proportion of scrambled scores ≥ observed score


Null Distribution of Alignment Scores

Adaptive Permutation P-Value

Increases permutations for sites with low p-values
to refine significance



Candidate Sites

10^3
Permutations

$p > 0.1$

Clearly not
significant –
stop testing

$p < 0.1$

10^4
permutations

$p > 0.001$

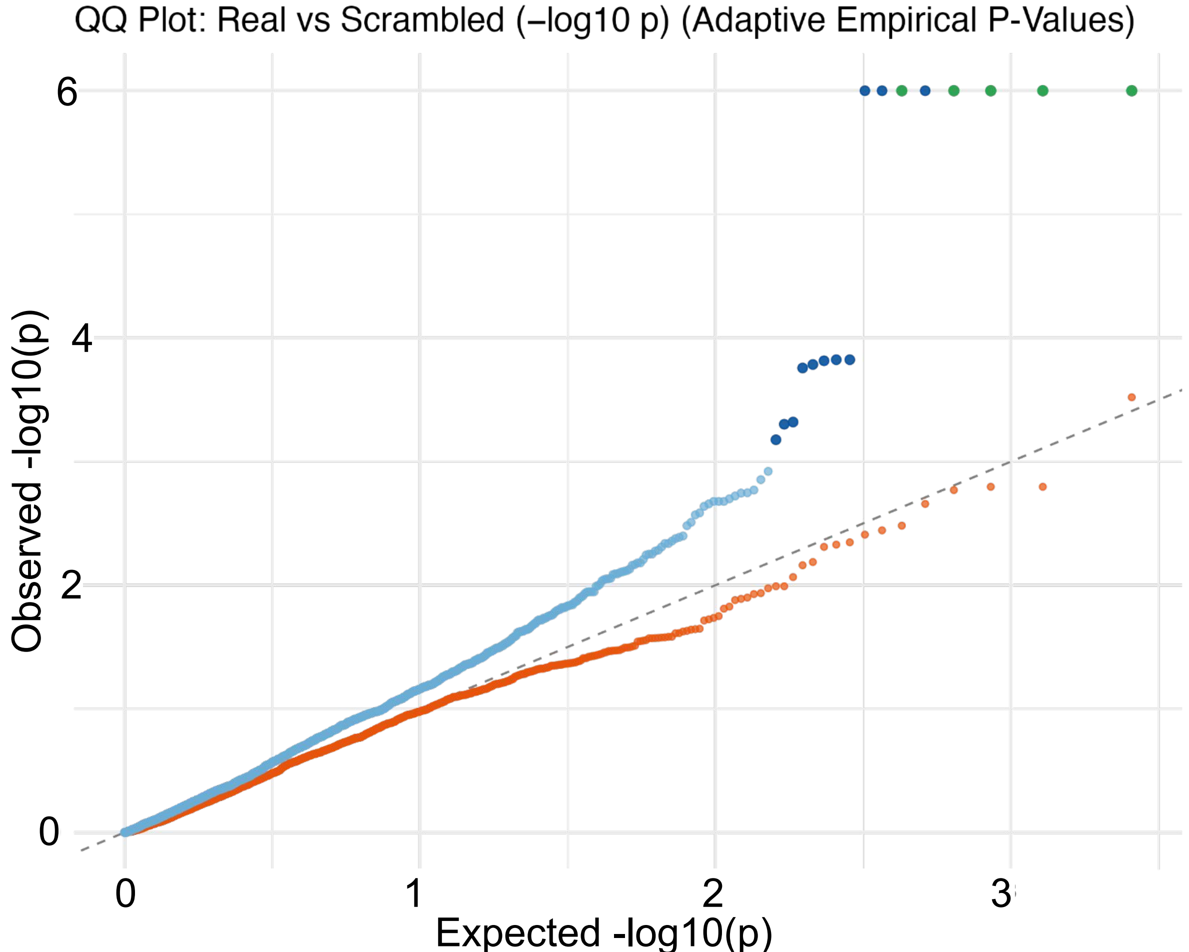
Not significant
– stop testing

$p < 0.001$

Continue
permuting

…

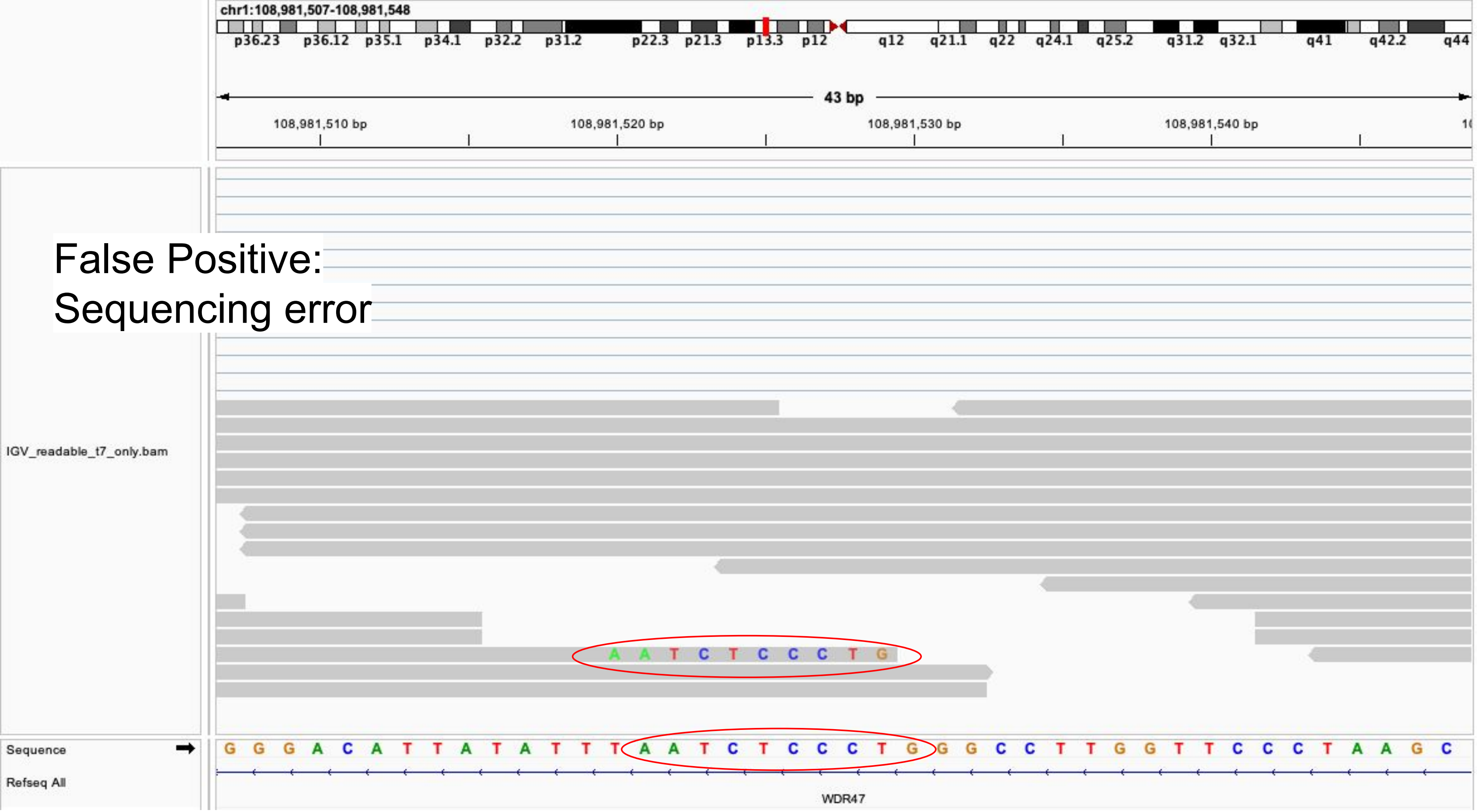
11 New Ultra-Rare Off-Targets Pass Statistical Thresholds


QQ Plot: Real vs Scrambled (–log10 p) (Adaptive Empirical P-Values)

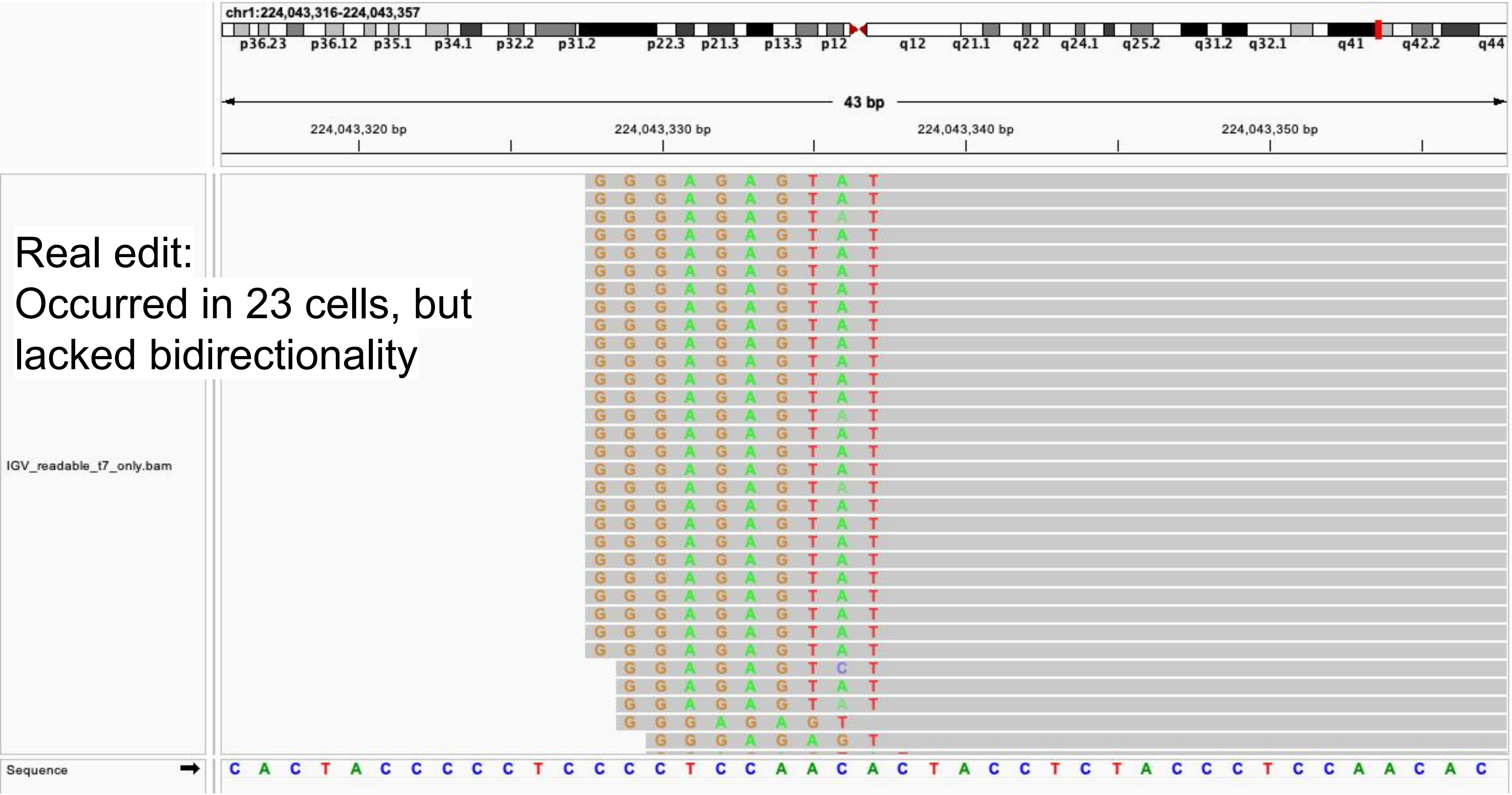
Significant Sites	Count
Confirmed (prior work)	5
New (this study)	11

UC San Diego

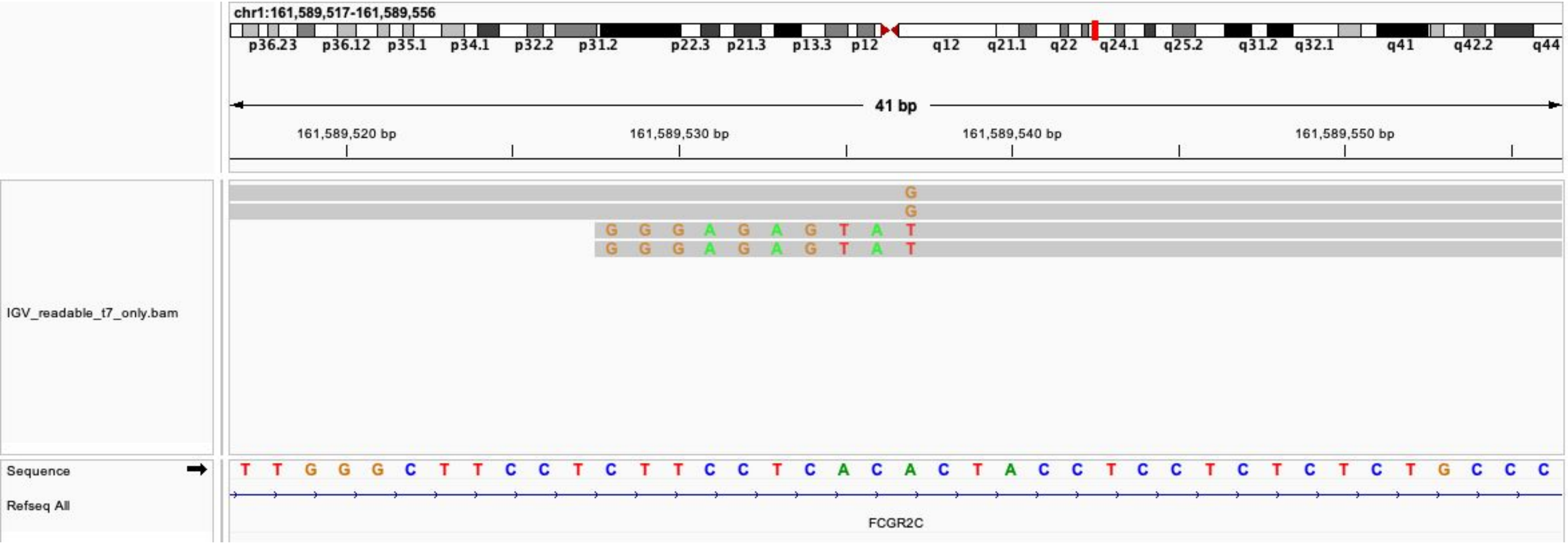
False Positive: Sequencing error



Real edit:
Occurred in 23 cells, but lacked bidirectionality



Real Edit:
Single cell edit event



Conclusions and Future Directions

Conclusions:

- Statistical analysis revealed 11 new significant candidate off-target sites
- 2 sites were manually validated as real edits missed by Sheriff
- Confirms existence of ultra-rare Cas9 off-target edits overlooked by standard Sheriff pipeline

Future Directions:

- Refine method to better detect rare edits and reduce false positives.
- Expand analysis to the rest of the genome and more guides
- Incorporate gene expression to assess functional impact of edits

Acknowledgements

- **Dr. Brad Balderson** – for mentorship, guidance on the Sheriff pipeline, and troubleshooting
- **Dr. Graham McVicker** – for project direction, feedback, and lab support
- **McVicker Lab Members** – for helpful discussions and advice throughout
- **UC LEADS/STARS Program** – for funding and supporting this summer research opportunity
- **UC San Diego Division of Biological Sciences** – for hosting the Summer Research Conference
- **You (the audience)** – for your time and attention

Thank you!
Any questions?

UC San Diego