CRISPR-Cas9 genome editing provides therapeutic opportunities, but unintended off-target edits remain a major concern. The McVicker Lab's Superb-seq assay jointly profiles single-cell genome edits and transcriptomes, but its standard pipeline misses potentially meaningful ultra-rare events.

To address this, I developed a statistical framework to evaluate ultra-rare edits identified using relaxed filtering conditions, expanding the catalog of low-frequency candidate sites. To rigorously distinguish true Cas9-mediated events from technical artifacts, I implemented a guide homology scoring pipeline . This pipeline performs global sequence alignment between each candidate site and known guide RNAs, assigning an alignment score based on sequence similarity and PAM proximity to reflect the likelihood of Cas9 targeting.

To assess statistical significance, I developed an adaptive empirical p-value method. This computationally efficient approach compares observed alignment scores against a null background of randomized guide-site pairs. The adaptive procedure dynamically increases permutations for initially significant candidates, providing high statistical resolution for promising sites while minimizing computational overhead.

This analysis successfully identified 11 new significant off-target sites on chromosome 1, all of which were missed by the default pipeline. Manual validation confirmed at least two of these as genuine, ultra-rare Cas9 edits, and the full set of significant sites showed statistical enrichment in intronic and promoter regions. This framework enhances Superb-seq's sensitivity to detect ultra-rare CRISPR-Cas9 off-target edits and provides a more rigorous statistical foundation for distinguishing true events from background noise, contributing to safer and more precise genome editing analyses.