**Bioinformatics analysis of SIKI sequencing data**

Sequenced libraries were first demultiplexed with Phenix [1]. The demultiplexed samples then underwent several processing steps, including preprocessing, trimming (Cutadapt [2]), UMI extraction (UMI-tools [3]), and UMI collapsing via the sikipipe pipeline (https://github.com/hukai916/sikipipe), an in-house Nextflow-based tool. For preprocessing, reads mapped to the minus strand were reverse complemented. Reads with lengths exceeding 1000 – 5000bp or lacking the expected universal sequences were considered problematic and were excluded from further downstream analysis.

Following UMI collapsing with sikipipe, the processed reads were analyzed using custom Python scripts integrated into another Nextflow pipeline called sikiclass (https://github.com/hukai916/sikiclass). This pipeline quantified the editing outcomes and categorized the reads into several groups: reads with precise tag inserts, reads with multiple tag inserts, and reads without tag inserts. Statistical tables were generated to summarize the ratios of each read category, the distribution of deletions and insertions for reads without tag inserts, and the SNP distribution for reads with precise tag inserts. The diagram below provides a detailed overview of the categorization process, and the criteria used.

A screenshot of a computer

Description automatically generated

1. Galanti L, Shasha D, Gunsalus KC: **Pheniqs 2.0: accurate, high-performance Bayesian decoding and confidence estimation for combinatorial barcode indexing.** *BMC Bioinformatics* 2021, **22:**359.

2. Martin M: **Cutadapt removes adapter sequences from high-throughput sequencing reads.** *EMBnetjournal*.

3. Smith T, Heger A, Sudbery I: **UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy.** *Genome Res* 2017, **27:**491-499.