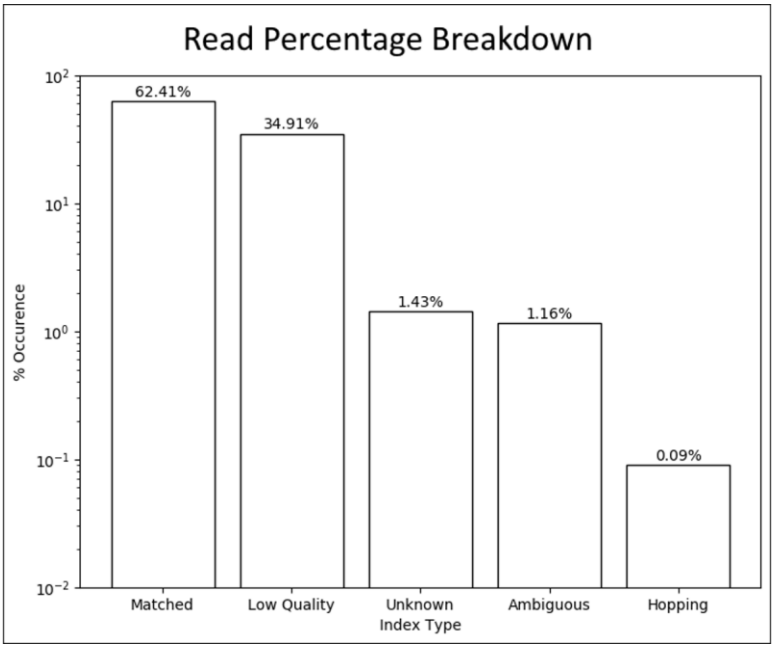

The Preceding chart displays a logarithmic breakdown of demultiplexing.py output files. Here is a description of each Index Type:

Index	**Description**
Matched	Reads with dual-matched indexes. As expected, most data fell in this category, indicating correct library preparation, sequencing, and demultiplexing.
Low Quality	Reads that fell below the quality-score cutoff. The proportion of data in this category is substantial, indicating the strict cutoff of 30 used in this demultiplexing.
Unknown	High-quality reads with 1+ non-ambiguous (no "N") indexes NOT present in the expected barcode list. These result from incorrect nucleotide assignment that Illumina claims is high-quality.
Ambiguous	Reads that contain 1+ "N" in their barcode, making the barcode ambiguous and therefore unusable. Interestingly, these barcodes appeared to cluster at the beginning of the read files.
Hopping	High-quality, unambiguous reads with different R2 and R3 barcodes, indicating the presence of index hopping.

Demultiplexing Rational

Ambiguous Reads

Ambiguous reads were filtered for on a per-nucleotide basis for each barcode. If reads contain N's, we cannot determine what barcode it is supposed to be, therefore making it useless.

Low Quality Reads

The quality score cutoff of 30 was chosen based on empirical observation of R2 & R3 mean index quality, where each nucleotide position had average scores >30. A threshold of 30 ensures data has 99.9% call accuracy, which is great.