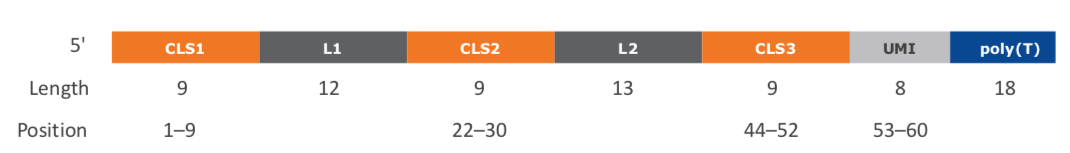
**Rhapsody R1 structure**



**Cell label**

Information on the cell label is captured by bases in three sections (CLS1, CLS2, CLS3) along each R1 read. Two common sequences (L1, L2) separate the three CLSs, and the presence of L1 and L2 relates to the way the capture oligonucleotide probes on the beads are constructed. By design, each CLS has one of 96 predefined sequences, which has a Hamming distance of at least four bases and an edit distance of at least two bases apart. A cell label is defined by the unique combination of predefined sequences in the three CLSs. Thus, the maximum possible number of cell labels is 963 (884,736). A cell label is represented by an index between 1–963.

Reads are first checked for perfect matches in all three pre-designed CLS sequences at the expected locations, CLS1: position 1–9, CLS2: position 22–30, and CLS3: position 44–52. Reads with perfect matches are kept.

The remaining reads are subjected to another round of filtering to recover reads with base substitutions, insertions, deletions caused by sequencing errors, PCR errors, or errors in oligonucleotide synthesis.

**UMI**

By design, the UMI is a string of eight randomers immediately downstream of CLS3. If the CLSs have perfect matches or base substitutions, the UMI sequence is at position 53–60. For reads with insertions or deletions within the CLSs, the UMI sequence is eight bases immediately following the end of the identified CLS3.

(If you want to compare directly cell by cell with our results)

**Conversion of 3 part cell label to single integer Cell\_Index**

After the list integer location of each cell label section is identified, the three part cell label is converted into a single integer with the following calculation:

(CLS1 - 1) \* 96 \* 96 + (CLS2 - 1) \* 96 + CLS3

Examples:

(First cell label in list from each section)

GTCGCTATA ACTGGCCTGCGA TACAGGATA GGTAGCGGTGACA AAGCCTTCT ACTGACTG

CLS1 1 L1 CLS2 1 L2 CLS3 1 UMI

Cell\_Index = (1-1) \* 96 \* 96 + (1-1) \* 96 + 1 = **1**

--------

(5th cell label in list from each section)

CGGTCCAGG ACTGGCCTGCGA CACCCAAAG GGTAGCGGTGACA GAACGACAA ACTGACTG

CLS1 5 L1 CLS2 5 L2 CLS3 5 UMI

Cell\_Index = (5-1) \* 96 \* 96 + (5-1) \* 96 + 5 = **37253**

--------

(96th cell label in list from each section)

TGCGATCTA ACTGGCCTGCGA CAACAACGG GGTAGCGGTGACA CATAGGTCA ACTGACTG

CLS1 96 L1 CLS2 96 L2 CLS3 96 UMI

Cell\_Index = (96-1) \* 96 \* 96 + (96-1) \* 96 + 96 = **884736**

--------

(43-12-77)

CTATGAAAT ACTGGCCTGCGA AAGCTACTT GGTAGCGGTGACA AACAGAAAC ACTGACTG

CLS1 43 L1 CLS2 12 L2 CLS3 77 UMI

Cell\_Index = (43-1) \* 96 \* 96 + (12-1) \* 96 + 77 = **388205**