- The *Cell Barcode* is similar to the sample barcode but there is (normally) no control over the assignment of cells to barcodes (whose sequence could be random or predetermined). The Cell Barcode can help identify when reads come from different cells in a "single-cell" sequencing experiment.

- The *UMI* is intended to identify the (single- or double-stranded) molecule at the time that the barcode was introduced. This can be used to inform duplicate marking and make consensus calling in ultra-deep sequencing. ~~Additionally, the UMI can be used to (informatically) link reads that were generated from the same long molecule, enabling long-range phasing and better informed mapping.~~ In some experimental setups opposite strands of the same double-stranded DNA molecule get related barcodes ~~.~~to differentiate from which strand of the double-stranded DNA molecule each read was observed. In this case, the
t MI tag can store not only the unique molecular identifier but also group reads that observe the top and bottom genomic strands respectively. These templates can also be considered duplicates even though technically they may have different UMIs. ~~Multiple~~ Additionally, the UMI can be used to (informatically) link reads that were generated from the same long molecule, enabling long-range phasing and better informed mapping. Finally, multiple UMIs can be added by a protocol, possibly at different time-points, which means that specific knowledge of the protocol may be needed in order to analyze the resulting data correctly.

BC:Z:`sequence` Barcode sequence (Identifying the sample/library), with any quality scores (optionally) stored in the `QT` tag. The `BC` tag should match the `QT` tag in length. In the case of multiple unique molecular identifiers (e.g., one on each end of the template) the recommended implementation concatenates all the barcodes and places a hyphen ('`-`') between the barcodes from the same template.

QT:Z:`qualities` Phred quality of the sample barcode sequence in the `BC` tag. Same encoding as QUAL, i.e., Phred score + 33. In the case of multiple unique molecular identifiers (e.g., one on each end of the template) the recommended implementation concatenates all the quality strings with spaces ('␣') between the different strings from the same template.

CB:Z:`str` Cell identifier, consisting of the optionally-corrected cellular barcode sequence and an optional suffix. The sequence part is similar to the `CR` tag, but may have had sequencing errors etc corrected. This may be followed by a suffix consisting of a hyphen ('`-`') and one or more alphanumeric characters to form an identifier. In the case of the cellular barcode (`CR`) being based on multiple barcode sequences the recommended implementation concatenates all the (corrected or uncorrected) barcodes with a hyphen ('`-`') between the different barcodes. Sequencing errors etc aside, all reads from a single cell are expected to have the same `CB` tag.

CR:Z:`sequence+` Cellular barcode. The uncorrected sequence bases of the cellular barcode as reported by the sequencing machine, with the corresponding base quality scores (optionally) stored in `CY`. Sequencing errors etc aside, all reads with the same `CR` tag likely derive from the same cell. In the case of the cellular barcode being based on multiple barcode sequences the recommended implementation concatenates all the barcodes with a hyphen ('`-`') between the different barcodes.

CY:Z:`qualities+` Phred quality of the cellular barcode sequence in the `CR` tag. Same encoding as QUAL, i.e., Phred score + 33. The lengths of the `CY` and `CR` tags must match. In the case of the cellular barcode being based on multiple barcode sequences the recommended implementation concatenates all the quality strings with with spaces ('␣') between the different strings.

MI:Z:`str` Molecular Identifier. A unique ID within the SAM file for the source molecule from which this read is derived. All reads with the same `MI` tag represent the group of reads derived from the same source molecule.

The MI tag value may end with a /[/] suffix indicating that it is one of several related barcodes[2].

[2]For example, `MI:Z:mol1/A` and `MI:Z:mol1/B` could be used to identify read pairs from the opposite strands of a duplex source molecule, where the `MI:Z:mol1/A` are by convention the "top (genomic) strand" reads and have 5' unclipped position of read one (of the pair) less than or equal to the 5' unclipped position of read two (of the pair). Then tools can find either the group of reads derived from that source molecule (those with the trimmed MI value `mol1`) or the groups of reads derived from each strand of that duplex source molecule (those with the full MI value `mol1/A`, or `mol1/B` respectively).