



- [1] Category of metadata (e.g. Diagnosis, Biospecimen, scRNA-seq Level 1, etc.); provide the same one for all items/rows.
- [2] Name of a file
- [3] Format of a file (e.g. txt, csv, fastq, bam, etc.)
- [4] HTAN Biospecimen Identifier (eg HTANx_yyy_zzz) indicating the biospecimen(s) from which these files were derived; multiple parent biospecimen should be comma-separated
- [5] Self-identifier for this data file - HTAN ID of this file HTAN ID SOP (eg HTANx_yyy_zzz)
- [6] The source of the input nucleic molecule
- [7] Indicate if library preparation was based on revived frozen cells.
- [8] The method by which cells are isolated into individual reaction containers at a single cell resolution (e.g. wells, micro-droplets)
- [9] The tissue dissociation method used for scRNASeq or scATAC-seq assays
- [10] Process which results in the creation of a library from fragments of DNA using cloning vectors or oligonucleotides with the role of adaptors [OBI_0000711]
- [11] Indicate if this is Read 1 (R1), Read 2 (R2), Index Reads (I1), or Other
- [12] Read 1 content description
- [13] Read 2 content description
- [14] The end of the cDNA molecule that is preferentially sequenced, e.g. 3/5 prime tag/end or the full length transcript
- [15] An oligo to which new deoxyribonucleotides can be added by DNA polymerase [SO_0000112]. The type of primer used for reverse transcription, e.g. oligo-dT or random primer. This allows users to identify content of the cDNA library input e.g. enriched for mRNA
- [16] A set of known synthetic RNA molecules with known sequence that are added to the cell lysis mix
- [17] The final concentration or dilution (for commercial sets) of the spike in mix [PMID:21816910]
- [18] A platform is an object aggregate that is the set of instruments and software needed to perform a process [OBI_0000050]. Specific model of the sequencing instrument.
- [19] Number of cells loaded/placed on plates
- [20] Number of cells and number of nuclei input; entry format: number, number
- [21] Number of days between sample for assay was received in lab and the libraries were prepared for sequencing [number]. If not applicable please enter 'Not Applicable'
- [22] Number of days between sample for single cell assay was received in lab and when the sample was dissociated and cells were isolated [number]. If not applicable please enter 'Not Applicable'
- [23] Number of days between sample for assay was received in lab and day of sequencing library construction [number]. If not applicable please enter 'Not Applicable'

[24] Number of days between sample for single cell assay was received in lab and day of nucleic acid capture part of library construction (in number of days since sample received in lab) [number]. If not applicable please enter 'Not Applicable'

[25] Protocols.io ID or DOI link to a free/open protocol resource describing in detail the assay protocol (e.g. surface markers used in Smart-seq, dissociation duration, lot/batch numbers for key reagents such as primers, sequencing reagent kits, etc.) or the protocol by which the sample was obtained or generated.

[26] A common term for all files belonging to the same cell or library. Provide a numbering of each library prep batch (can differ from encapsulation and sequencing batch)

[27] Unique cell barcode assigned to empty cells used as controls in CEL-seq2 assays.

[28] Indicate if protein expression (EPCAM/CD45) positive/negative data is available for each cell in CEL-seq2 assays

[29] Unique ID for this feature. Must not contain whitespace, quote or comma characters. Each ID must be unique and must not collide with a gene identifier from the transcriptome [<https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/using/feature-bc-analysis#feature-ref>]

[30] Sample Level HTAN Data File ID for the associated level - HTAN ID of this file HTAN ID SOP (eg HTANx_yyy_zzz)

[31] The method by which cells are multiplex or labeled with cell surface markers or probes

[32] The library construction methods for the feature barcode library

[33] Synapse ID of the folder containing the barcode lists

[34] A comma separated list of filenames in the gzipped folder detailing what barcodes are specific to demultiplexing samples versus providing surface protein data

[35] Length of cell barcode read (in bp): number

[36] Offset in sequence for cell barcode read (in bp): number

[37] Number

[38] Length of UMI barcode read (in bp): number

[39] Start position of UMI barcode in the sequence. Values: number, 0 for start of read

[40] Number

[41] Length of cDNA read (in bp): number

[42] Offset in sequence for cDNA read (in bp): number