

Metadata Specifications:

Cell Lines

Immortalized (naturally or engineered), genetically uniform tissue cells able to reproduce indefinitely in standard culture conditions.

Importance	1: Required, 2: Required if available, 3: Optional
Common Fields	Fields that are common across all LINC Standards
Custom Fields	Fields that are unique to a single LINC Standard or common across only a subset of them

Common Fields

LINC Field Name	Related to	Description	Comments	Importance
CL_LINCS_ID	Canonical	Unique LINC internal identifier	LINC internal ID; this is a batch independent ID; canonical cell line ID	1
CL_Name	Canonical	The primary name for the cell line as chosen by LINC	Should be descriptive and correspond to existing cell line names as much as possible; batch independent name	1
CL_Alternative_Name	Canonical	Other relevant names	Synonymous or alternative names; but only significantly different names should be captured	2
CL_Alternative_ID	Canonical	This field specifies the CLO, if available, or other common ID for the cell line or, if derived from another line, the CLO or other ID for the parent cell line. The parent cell line ID must be propagated to all cell lines derived from that parent line unless a distinct CLO has been assigned to the derived line	CLO or other common IDs referring to the same cell line	2
CL_Center_Canonical_ID	Canonical	LINC DSGC-specific canonical ID. This will be assigned by a given LINC DSGC according to its cell line registration scheme	-	1
CL_Relevant_Citations	Canonical	List of references (with PMIDs) of relevance to cell line derivation, etc.	-	2
CL_Center_Name	Batch	Name of the LINC Center that is using the Cell Line	-	1
CL_Center_Batch_ID	Batch	LINC center-specific cell line ID; batch specific ID	LINC DSGC-specific cell line batch ID. This will be assigned by a given LINC center according to its cell line registration scheme	1
CL_Provider_Name	Batch	Name of vendor or lab (provider) that supplied the cell line	ATCC or other vendor(s) or provider	1
CL_Provider_Catalog_ID	Batch	ID or catalogue number or name assigned to the cell line by the vendor or provider	ATCC or other cell line provider's IDs	1
CL_Provider_Batch_ID	Batch	Vendor/Provider Batch ID number; Batch or lot number assigned to the cell line by the vendor or provider	provided by the cell line provider	1
CL_Comments	Batch	DSGC Comments regarding reagent	-	3

Custom Fields

CL_Organism	Canonical	Organism of origin; a controlled vocabulary describing the organism from which the cell line was derived (e.g. Homo sapiens, Mus musculus, etc.)	exact NCBI name	1
CL_Organ	Canonical	Organ of origin; controlled terms describing the organ from which cell line is derived; (e.g. lung, mammary gland etc.)	-	1
CL_Tissue	Canonical	Tissue of origin; A controlled vocabulary describing the tissue from which the cell line was derived	Some histology information might be provided in this field.	1
CL_Cell_Type	Canonical	A controlled vocabulary describing the cell type from which a cell line was derived; e.g. epithelial like, fibroblast-like, lymphoblast like, hematopoietic, mesenchymal, neural, etc. This provides information about cell morphology. Also sometimes referred to as cell morphology	controlled terminology from CL	1
CL_Cell_Type_Detail	Canonical	Additional description of cell type (histology) that is not available in CL, but may be known from other sources like ATCC	terms from other sources like ATCC; will develop over time	2
CL_Donor_Sex	Canonical	Describes sex of the organism from which the cell was obtained	male, female, or genderless; OBI	2
CL_Donor_Age	Canonical	The age of the donor	numeric number; donor age in years	2
CL_Donor_Ethnicity	Canonical	For human cells, the ethnicity of the donor	-	2
CL_Donor_Health_Status	Canonical	Controlled vocabulary describing the health status of the donor	need to be defined in more detail; need level of detail required	2
CL_Disease	Canonical	If the cell line came from a particular diseased tissue, the disease should be noted in terms of a controlled vocabulary (e.g. breast cancer, colon cancer, not diseased, etc.)	the disease hierarchy is captured in the ontology; i.e. DOID	1
CL_Disease_Detail	Canonical	Additional description of a disease related to the cell line that may not be available in the disease ontology above	need to develop what exactly should go here and the corresponding terms	2
CL_Known_Mutations	Canonical	Mutations inherent (from the donor) in the cell line, captured explicitly; e.g. if reference is not available	Needs some ontology to describe gene / protein and mutation; at this point we suggest a concatenation of UniProt / Gene symbol and code of mutation	2
CL_Mutation_Citations	Canonical	Mutations inherent in certain cell lines; from a reference	Known mutation in cell line from a reference; needs to include the reference source and the reference to the specific cell	2
CL_Molecular_Features	Canonical	Relevant molecular and morphological features of the Cell Line	e.g. ER Status, Luminal Cells	3
CL_Genetic_Modification	Canonical	This field specifies any stable constructs as well as any genetic modifications (mutations, translocations) introduced into this cell line (e.g. H2B-mCherry integrated at the AAVS1 Safe Harbor locus). Details of the procedures used to generate this line (e.g. CRISPR/Cas9-mediated transformation) should be described and appropriate citations provided in the Production Details field	MIACA is minimal information that may be a guidance; requires more fields to define modifications using controlled terms	1
CL_Growth_Properties	Canonical	A controlled vocabulary describing the growth properties of the cell line (e.g. adherent, suspension)	-	1
CL_Recommended_Culture_Conditions	Canonical	A description of the standard tissue culture conditions (media, supplements, culture dish treatment) used to maintain the cell line. Description of culture dish treatment conditions would include information about coating of culture dish with fibronectin, collagen, etc. prior to cell plating. If special culture vessels are required to grow the cells, these should also be mentioned and details provided	Recommended standard culturing conditions go here; not a required field; the actual culture conditions are captured as experimental metadata	2
CL_Related_Projects	Canonical	Other projects in which the cell line has been studied / used; A controlled vocabulary describing other large scale projects in which the cell line has been used (e.g. ENCODE, TCGA, ICBP, Epigenomics, etc.)	Needs defined project codes	3
CL_Verification_Reference_Profile	Canonical	expected STR (reference) profile of the cell line based on provider information, if available	from cell line provider / reference	2
CL_Reference_Source	Canonical	This field specifies a catalog entry or database record for this cell line at a cell type collection, if available, or an established external source	e.g. ATCC or RIKEN, or Joe Smith (University of MI)	1
CL_Cell_Markers	Canonical	A controlled vocabulary describing the markers used to isolate / identify the cell type	controlled terms of markers; at this point no reference	2
CL_Gonosome_Code	Canonical	List of the sex chromosomes (gonosome) of the sample e.g. XX, XY, XXY	-	3
CL_Disease_Site_Onset	Canonical	Site of disease onset in primary cell donor	Primary Cell / Cell Line of Origin Information	3
CL_Disease_Age_Onset	Canonical	Age of disease onset in primary cell donor (in years)	Primary Cell / Cell Line of Origin Information	3
CL_Donor_Age_Death	Canonical	Age of death of primary cell donor (in years)	Primary Cell / Cell Line of Origin Information	3
CL_Donor_Disease_Duration	Canonical	Disease duration in cell donor: Age of Sample Acquisition - Age of Onset, (in years)	Primary Cell / Cell Line of Origin Information	3
CL_Precursor_Cell_Name	Canonical	This field specifies the name of the parent cell line from which the cell line was derived. It is left blank if this cell line was not known to be derived from another. The particular batch of the parent line that was used to generate the new line should be specified in the Production Details (canonical) field	-	1
CL_Precursor_Cell_LINCS_ID	Canonical	This field specifies the global LINC ID of the parent cell line from which the cell line was derived. It is left blank if this cell line was not known to be derived from another. The particular batch of the parent line that was used to generate the new line should be specified in the Production Details (canonical) field.	-	1
CL_Precursor_Cell_Center_Batch_ID	Canonical	If this cell line is derived from another registered cell line, this field should specify the Center LINC Batch ID of the specific batch from which it was derived	-	2
CL_Production_Details	Canonical	This field specifies the procedure(s) by which the cells were derived from the parent/precursor cell, including genetic transformations and phenotypic selections. Citations / source information for constructs and citations for procedures should be included here when appropriate	-	2
CL_Quality_Verification	Batch	Information pertaining to experimental verification of the cell line identity; batch-specific ID; STR profiling	Acceptable protocols for verification will be determined by LINC participants and a controlled vocabulary will be developed	2
CL_Transient_Modification	Batch	Transient transfection or viral transduction	need to capture transfection agent	1
CL_Passage_Number	Batch	The number of times, if any that the cells have been re-plated and allowed to grow back to confluency or to some maximum density if using suspension cultures	-	2
CL_Source_Information	Batch	This is a free-text field that provides detailed source information for this particular batch, which may include information on from whom and when the provider obtained the cells and for what purpose the cells were obtained by the end user	-	2
CL_Date_Received	Batch	This field specifies when this batch was obtained from the provider. Because YEAR-MO-DY is not always known, this field may only contain partial date information (e.g. YEAR-MO, YEAR only)	-	2
CL_Center_Specific_Code	Batch	This field specifies the commonly used code or name for this batch at this Center, which may differ from the Provider Catalog ID and from the canonical cell line name	-	3