

CMO Patient ID and Sample ID generation

Purpose of CMO Sample ID is to provide a unique, human readable identifier encoding patient ID and specimen type, to be utilized in CMO processing

- CMO Sample ID is generated while sample is promoted on Sample Intake website.
- CMO Sample ID is generated only for samples with species "Human" or samples with recipe listed in file (even if species is NOT "Human" or is empty):
 - igo.mskcc.org: /srv/www/sapio/tomcat7/webapps/LimsRest##-<VERSION>/WEB-INF/classes/app.properties
 - tango.mskcc.org: /srv/www/sapio/tomcat7/webapps/LimsRest##-<VERSION>/WEB-INF/classes/app.properties
 - in property "human.recipes", eg.:

human.recipes=IMPACT341,IMPACT410,IMPACT410+,IMPACT468,IMPACT505,HemePACT_v3,

HemePACT_v4,MSK-ACCESS_v1

- CMO Sample ID is stored in LIMS in Sample Level Info Record in column CorrectedCMOID (Corrected CMO ID).
- CMO Sample ID may be regenerated during running workflow "Adding or Amending CMO Information" in LIMS if changes result in different CMO Sample ID.

cBioPortal - <https://one.mskcc.org/sites/pub/corp/GenoEffort/Pages/cBioPortal.aspx>

There are two main types of CMO Sample ID: Cell Line and C space (patient oriented) id:

Cell Line CMO Sample Id

Used if sample's Specimen Type (stored in LIMS in Banked Sample record in column SpecimenType (CMO Sample Type)) is equal to "CellLine" AND NormalizedPatientId (Normalized Patient Id) is different than "MRN_REDACTED".

Format: <Investigator Sample Id>-<Request Id>, eg. JH1234-12345P

- **Investigator Sample Id** - sample id provided by investigator in Rex (Sample Intake).
- **Request Id** - request to which a sample belongs (request id is assigned during sample promotion).
 - underscores are removed from Request id in CMO Sample ID, eg. if sample id equals to JH123 and request id equals to 12345_T - CMO Sample Id will be equal to: JH123-12345T
 - LIMS/rex site adds next request 'letter' in order during sample promotion.

C space (patient oriented) CMO Sample Id

Used if sample's Specimen Type is **not** equal to "CellLine" OR NormalizedPatientId (Normalized Patient Id) is equal to "MRN_REDACTED".

Format: <Patient Id>-<Sample Type Abbreviation><Sample Counter>-<Nucleid Acid Abbreviation><Nucleic Acid Counter>, eg: C-123456-X001-d01¹

- **Patient Id** - id generated by CRDB, stored in LIMS in Banked Sample Record in column CMO Patient ID (CMO Patient ID, which goes to the same field in Sample Level CMO Information record)

¹ The <Nucleic Acid Counter> was added to the CMO sample label in January 2022. Labels generated prior to January 2022, either by IGO or by SMILE, did not include this counter and simply ended with the <Nucleic Acid Abbreviation>

- **Sample Type Abbreviation** - uppercase one letter long abbreviation from one of the possible values: {T, R, P, M, N, L, U, S, X, G, A, F}²³
 - if sample **Specimen Type** (stored in LIMS in Banked Sample record in column SpecimenType (CMO Sample Type)) equals to one of those values it's mapped accordingly:
 - "PDX" => "X"
 - "Xenograft" => "X"
 - "XenograftDerivedCellLine" => "X"
 - "Organoid" => "G"
 - if sample Specimen Type is equal to "cfDNA" - Sample Type Abbreviation is resolved using **Sample Origin** (stored in LIMS in Banked Sample record in column SampleOrigin (Sample Origin)):
 - "Urine" => "U"
 - "Cerebrospinal Fluid" => "S"
 - "Plasma" => "L"
 - "Whole Blood" => "L"
 - if Specimen Type is equal to "Exosome" - Sample Type Abbreviation is resolved using **Sample Origin** (stored in LIMS in Banked Sample record in column SampleOrigin (Sample Origin)):
 - "Plasma" => "L"
 - "Whole Blood" => "L"
 - Else, default to "Unknown Tumor" => "T"
 - if Specimen type is neither of above values - **Sample Class** is used to resolve Sample Type Abbreviation (stored in LIMS in Banked Sample record in column Sample Class (SampleClass)):
 - "Unknown Tumor" => "T"
 - "Local Recurrence" => "T"
 - "Primary" => "T"
 - "Recurrence" => "T"
 - "Metastasis" => "T"
 - "Normal" => "N"
 - "Adjacent Normal" => "N"
 - "Adjacent Tissue" => "A"
- **Sample Counter** - integer value incremented for each patient's sample with the same Sample Type Abbreviation
 - 3 digit integer value from range 001-999 (values less < 100 are filled in with zeros '0' to preserve 3-digit format)
 - first sample for particular patient with particular Sample Type Abbreviation gets value 001
 - counter is shared for all patient's samples belonging to all projects (eg. if in request 12345_A there is a sample C-123456-T001-d for patient C-123456 and new CMO Sample Id is generated for another "T" sample for request 12345_B - CMO Sample Id will have value: C-123456-T002-d)
 - Note: This assumes that these two samples have different ALT IDs.
 - If a sample specimen is resequenced under a new IGO ID, the existing <Sample Type Abbreviation><Sample Counter> will be used for the new CMO label generated. Instead of incrementing the Sample Counter, the Nucleic Acid Counter will be incremented.
 - Note: assuming that no metadata corrections were made that would result in a different Sample Type Abbreviation, such as a switch from Tumor to Normal. If this is the case then the correct sample type designation needs to be identified and the metadata should be updated for samples to reflect the correct value.
- **Nucleic Acid Abbreviation** - lower case one letter long abbreviation from one of the possible values: {d, r}
 - Nucleic Acid Abbreviation is inferred from Banked Sample Record from column SampleType (this column is filled in from the sequencing assay request chosen on Sample Intake website in field "I want to send to IGO")
 - "Pooled Library":
 - if Recipe is "RNASeq" "r"
 - if Recipe is anything else than "RNASeq" "d"
 - "DNA" / "cfDNA" / "DNA Library" "d"

² Sample Type Abbreviations "A" and "F" were added after SMILE took over the CMO label generation. The abbreviation "F" is a special abbreviation that was added to catch samples that would otherwise fail to get a label generated but still allow the sample to be published and processed by downstream consumers, like Voyager. In other words, "F" was added as a way to prevent samples from being held back from processing simply due to not having a CMO label generated.

³ Sample Type Abbreviations designating "Tumor" samples are consolidated/simplified to "T". Historically, "Tumor" designations were also indicated by "P", "R", and "M".

- "RNA" "r"
 - other Nucleic Acid Abbreviation is retrieved from LIMS from Banked Sample Record from column NAtoExtract (Nucleic Acid Type to Extract)
 - "DNA" "d"
 - "DNA and RNA" "d"
 - "RNA" "r"
 - "cfDNA" "d"
- **Nucleic Acid Counter** - integer value incremented for each patient's sample with the same Sample Type Abbreviation, ALT ID, and Nucleic Acid Abbreviation.
 - 2 digit integer value from range 01-99 (values less < 10 are filled in with zeros '0' to preserve 2-digit format)
 - first sample for particular patient with particular Sample Type Abbreviation, ALT ID, and Nucleic Acid Abbreviation gets value 01
 - Samples with labels generated prior to January 2022 will not have a Nucleic Acid Counter. These samples are considered to have “legacy” IGO labels and their Nucleic Acid Counter is treated as 01 when determining what Nucleic Acid Counter subsequent patient samples should be assigned.
 - Example: A second sample for a particular patient with a particular Sample Type Abbreviation, ALT ID, and Nucleic Acid Abbreviation gets value 02, and so on.

CMO Patient ID

CMO Patient ID will be dynamically generated/linked by CRDB on submission, and the CRDB service can optionally be called again by an SM/PM before Promotion from BankedSample to "real" sample in LIMS.

There are three cases:

- The submitter used an MRN as patient ID – CRDB will return the anonymous patient ID for this MRN. We will store the constant string "MRN" in submitted ID in banked sample, as well as the CMO Patient ID.
- The submitter used a string as patient ID and marked the sampletype as CELL LINE.
 - We will attempt to normalize the cell line by converting to all upper or lowercase and removing spaces hyphens and underscores.
 - We will send the normalized string to CRDB and again receive the anonymized CRDB ID back. In cases where normalization is performed, we will store the prenormalized string in BankedSample, as well.
- The submitter used a string and the sample type is not equal to CELL LINE.
 - We will send the CRDB the supplied patient ID + the "requestor ID" (which is the "user_name" of the sample submitter currently logged in to the Sample Intake website)

Service [CRDB CMO id service](#) is used to retrieve CMO Patient Id.

CMO Sample ID regeneration

If a change in LIMS done during running workflow "Adding or Amending CMO Information" results in different CMO Sample ID as currently exists, notification is shown informing about new CMO Sample ID.

If CMO Sample ID change is required, there is no possibility to change information in LIMS without regenerating CMO Sample ID. Either both will be changed or none of them.

User can cancel the workflow if does not accept the new CMO Sample ID but desired change to Sample Level Info record won't be saved as well.