



TOPMed Omics sample identifiers and documentation

Updated 08/16/2022

TOPMed sample identifiers

The TOPMed IRC will distribute sample identifiers for multiple types of Omics samples. These IDs will begin with three letters, followed by 6 integers. The three-letter prefix indicates the type of sample:

- TOR – TOPMed Omics RNA sample
- TOM – TOPMed Omics Metabolite sample
- TOE – TOPMed Omics DNA sample for epigenomic assay (e.g. methylation)
- TOP – TOPMed Omics sample for serum/plasma proteomic assay

These IDs are analogous to the “NWD” IDs used for DNA samples for WGS.

These IDs are intended to represent a unique analyte sample/assay instance identifier across all TOPMed studies. In the following, **we make a distinction between analyte sample ID and subject (individual) ID**. Note that where a single subject contributes multiple sample analyte aliquots (i.e. multiple body sites, replacements, duplicates), each of those aliquots should have a different analyte sample ID.

The file(s) resulting from an assay instance (e.g. RNAseq) will be labeled with the appropriate TOPMed sample identifier (e.g. TOR ID for RNAseq). The assay center will provide a sample manifest and 2D-barcoded tubes (or other container) for supplying the samples. The manifest will contain the 2D-barcode and an empty column for the TOPMed sample IDs which will be filled in by the study. You may also wish to include your local analyte sample ID on the manifest, if allowed by the assay center.

Study investigators are responsible for assigning TOPMed sample identifiers to subjects and for maintaining and providing sample-subject mapping files. These mapping files must include the subject ID that is used in phenotype files previously posted on dbGaP. If phenotype data have not previously been posted on dbGaP, please use subject IDs that are postable (i.e. de-identified). Study investigators are also responsible for recording and providing the provenance and attributes of the analyte samples (see documentation requirements below).

The contact email for obtaining new whole genome sequence or omics sample identifiers is: topmed.omics@umich.edu [1].

Sample submission procedure (e.g. RNA samples)

1. The IRC will send a block of TOR_IDs to each study investigator, with ~20% extra to cover replacements and other contingencies. Each TOR_ID consists of “TOR” followed by 6 integers (e.g. TOR103482).

2. Each assay center will send a sample manifest and 2D-barcoded tubes or plates to the study investigators. The manifest will contain columns for the barcode, the TOR_ID, the de-identified local sample ID normally used by the project, and other columns for sample annotation such as sex and ethnicity. (The local sample ID is for convenience in preparing sample submissions; it will not be used for dbGaP posting.)
3. Each study will fill out the sample manifest, which involves linking the barcode, the TOR_ID, and the local sample ID. Please discuss with your sequencing center contact how to indicate samples that are included as extras for replacement of sample failures.
4. **Each analyte sample aliquot gets a different TOR_ID. If you replace one sample with another from the same individual, the replacement aliquot should be assigned a different TOR_ID.**
5. The sample manifest will be checked by the assay center and, when approved, the study investigator will place samples into the appropriate barcoded tubes and ship to the sequencing center.
6. At each sequencing center, the TOR_ID will be linked to a LIMS ID. Only these two IDs will be propagated into the data files.
7. The study investigator will provide to the IRC all of the documentation requirements listed below. The final sample manifests will be collected by the IRC and maintained as a record of the ID linkages.

Documentation requirements

The TOPMed program requires that omics data be submitted to dbGaP, along with thorough documentation of biosampling and laboratory methods, as well as sample provenance. Studies will need to submit to TOPMed sample attributes files and protocol documents that describe the clinical and laboratory procedures used at each step in the process of supplying an analyte sample to the center performing the omics assay.

The documentation requirements are explained in the following two documents:

- [General instructions - Omics Metadata from TOPMed Studies](#) [2]
- [Data dictionary template](#) [3] (for sample attributes files)

These documents must be submitted to the dbGaP submission portal before omics data are transferred to the study.

For questions, please contact Topmed-Admin@westat.com [4] or topmed.omics@umich.edu [1].

Source URL (modified on 08/16/2022 -

12:28pm):<https://topmed.nhlbi.nih.gov/topmed-omics-sample-identifiers>

Links

[1] <mailto:topmed.omics@umich.edu> [2] <https://topmed.nhlbi.nih.gov/topmed-omics-metadata-requirements> [3]

<https://topmed.nhlbi.nih.gov/system/files/documents/TOPMed%20omics%20metadata%20dictionary.xlsx> [4]

<mailto:Topmed-Admin@westat.com>