Published on *NHLBI Trans-Omics for Precision Medicine* (https://topmed.nhlbi.nih.gov)

NWD_ID Allotment Instructions

Updated 12/21/2023

TOPMed DNA sample identifiers - NWD_IDs

The NWD_ID was created to be a unique DNA sample/sequencing instance identifier across all TOPMed studies. In the following, we make a distinction between DNA 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 DNA sample ID.

In TOPMed, each DNA sample will result in a different sequencing instance, recorded in a BAM file. The NWD_ID of that sample will be in the header of the resulting BAM file and these are the sample IDs that will be posted on dbGaP. The numbered points below summarize the process of sample submission to the sequencing centers with respect to the NWD sample ID.

When completed by you, the sequencing center sample manifest will provide a mapping among the 2D-barcode on the tube, the NWD sample ID and your local DNA sample ID. You will also need to provide a "SubjectSampleMapping" file, giving the correspondence between the sample IDs in the sequencing center manifest and the subject ID that you have used previously for dbGaP posting (or the subject ID you will use for dbGaP posting if that has not been done previously). This subject ID must be de-identified. Please refer to the dbGaP Study Submission Guide [1] for a description of the SubjectSampleMapping [2] and other similar files that must be submitted to dbGaP in order to register your TOPMed WGS study.

Sample submission procedure

1. Each NWD_ID consists of "NWD" followed by 6 integers (e.g. NWD103482), where "NWD" stands for NHLBI WGS DNA sample.
2. The DCC will send a block of NWD IDs to each study investigator, with ~20% extra to cover replacements and other contingencies
3. Each sequencing 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 NWD 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.)
4. Each study will fill out the sample manifest, which involves linking the barcode, the NWD_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.
5. **Each DNA sample aliquot gets a different NWD_ID. If you replace one DNA sample with another from the same individual, the replacement aliquot should be assigned a different**

NWD_ID.

6. The sample manifest will be checked by the sequencing center and, when approved, the study investigator will place DNA samples into the appropriate barcoded tubes and ship to the sequencing center.
7. At each sequencing center, the NWD ID will be linked to a LIMS ID and only these two IDs will be propagated into the data files.
8. The final sample manifests will be collected by the DCC and maintained as a record of the ID linkages.
9. The study investigator will also submit to dbGaP and the DCC a separate file of linkages between the two sample IDs (NWD and local) and subject IDs ("SubjectSampleMapping").

Links

[1] https://www.ncbi.nlm.nih.gov/gap/docs/submissionguide/ [2] https://www.ncbi.nlm.nih.gov/gap/docs/submissionguide/#10-how-do-i-create-subject-sampl