SMaHT Sample & File Nomenclature

Resource for the SMaHT Network Members
Version 2.1
Last modified: 2026-Jan-02

Description **1**
Overview **1**
- Naming schema 2
- Donors 2
  - Donor identifier for Tissues 2
  - Donor identifier for Benchmark Cell Lines 2
- Sample Types 3
  - Tissue 3
  - Cell Lines 8
- Sample and Assay Metadata in the File Name 8
- File and Other Information in the File Name 12

Examples **13**
- Benchmarking Cell Line Names 13
- File Names 14
  - Files Generated from Benchmarking Samples 14
  - Files Generated from Production Samples 15

Description

The SMaHT sample and file names are the primary identifiers of biosamples from the Tissue Procurement Center (TPC) and files generated by the Data Analysis Center (DAC) of the SMaHT Network ("Network").

Overview

The SMaHT sample and file names contain identifiers that are unique and immovable, as well as semi-human-readable codes that correspond to metadata. This document describes the naming schema and tables of codes for each metadata type that are included in sample and file names. The metadata fields in the sample and file names are delimited by a hyphen ("-"). "`#`" indicates a single-digit integer number, and "`A`" indicates an alphabetical letter in this document.

Naming schema

[Project][Kit/Donor ID] **-** [Protocol ID] **-** [Tissue aliquot & Tissue core IDs] **-** [Sex & Age] **-** [Sequencing platform & Sequencing assay code] **-** [Center code] **-** [Unique Accession ID] **-** [Analysis info] **.** [File format]

Donors

Donor identifier for Tissues

The donor identifiers are assigned by the Tissue Procurement Center at NDRI (TPC), following the schema below.

Schema: **ProjectDonorKitID**

Benchmarking donors: **ST###**

Production donors: **SMHT###**

Non-benchmarking, non-production donors: **SN###**

For the benchmarking donors, the project is "**ST**".

For the production donors, the project is "**SMHT**".

(Note: For some donors, tissues were provided by TPC for the Tools and Technology Development groups (TTDs) to test their protocols only, not related to the benchmarking nor production studies. For these non-benchmarking, non-production donors, the project is "**SN**").

- **Project**: **ST** for a benchmarking donor; **SMHT** for a production donor.
- **Kit ID**: Provided by TPC/NDRI. It is a three-digit pre-generated kit identifier that references the donor. `001-150` are reserved for production donors; and `001-004` for benchmarking donors.

Donor identifier for Benchmark Cell Lines

Some benchmarking samples are derived from well-established cell lines, such as COLO829 and HapMap cell lines. As these are not derived from the TPC donors, the donor identifiers are assigned shown in **Table 1**, preceded by the prefix, "**SMHT**", e.g., **SMHTCOLO829T**, **SMHTHAPMAP6**.

Table 1. Benchmarking cell line codes.

Kit/Sample ID	**Cell line description**
COLO829T	COLO829 tumor cell line

COLO829BL	COLO829BL normal lymphoblast cell line
COLO829BLT50	COLO829 1:50 admixture
HAPMAP6	Cell admixture of six HapMap cell lines
LBLA2	LB-LA2 fibroblast cell line
LBIPSC1	iPSC line from clone #1 derived from the LB-LA2 fibroblast cell line
LBIPSC2	iPSC line from clone #2 derived from the LB-LA2 fibroblast cell line
LBIPSC4	iPSC line from clone #4 derived from the LB-LA2 fibroblast cell line
LBIPSC52	iPSC line from clone #52 derived from the LB-LA2 fibroblast cell line
LBIPSC60	iPSC line from clone #60 derived from the LB-LA2 fibroblast cell line

Sample Types

Tissue

The metadata and identifiers of tissues and samples sent to sequencing centers/labs are generated at the TPC.

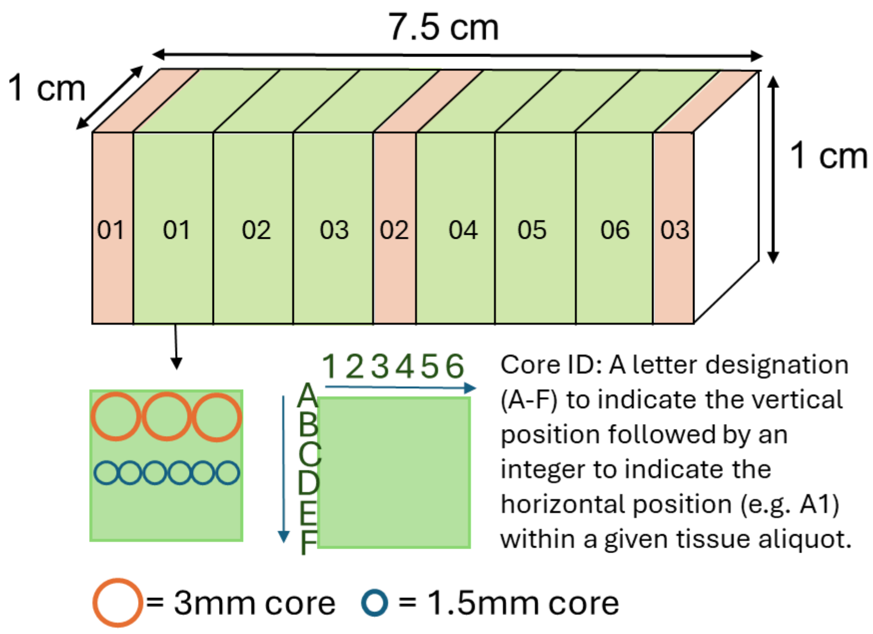
Schema: **ProjectDonorKitID – TissueProtocolID – TissueAliquotIDCoreID**

Benchmarking tissues: `ST### – #A – ###A#`
Production tissues: `SMHT### – #A – ###A#`
Non-benchmarking, non-production tissues: `SN### – #A – ###A#`

Note:

1. Unless noted, all benchmarking tissues are homogenized tissues. At specific requests by certain GCCs or TTDs, non-homogenized intact benchmarking tissues were made available for sequencing.
2. Intact tissues are used for all production tissues. They are *not* homogenized.
3. For intact tissues, solid tissues are referred to as “tissue aliquots”, which subsequently get sectioned into multiple “tissue cores” or labelled as a single “tissue specimen”, depending on the availability of the tissue material (tissue specimens are larger than cores and smaller than tissue aliquots).
4. GCCs do not rename the sample identifiers received from TPC and submit the metadata to DAC. GCCs provide DAC with the metadata they received from TPC without any alteration and provide additional metadata for any information related to further sample processing done at GCC.

- **Protocol ID**: Automatically generated by NDRI’s Rhythm system. A single internal protocol group number is followed by a letter designation (`A-ZZ`) corresponding to a specific combination of tissue type and tissue preservation method (**Table 2**).
- **TissueAliquot IDs:** Provided by TPC/NDRI. An aliquot can refer to either a solid tissue aliquot or a liquid tissue aliquot in a vial (e.g. whole blood). An aliquot ID is a three-digit number (e.g., “001”).
    - For solid tissues:
        - Aliquot ID refers to spatially numbered solid tissue aliquots from longitudinally bisected or non-bisected tissue samples. Serial tissue cores subsampled from within an aliquot are designated by a letter followed by a number (Figures 1 and 2).
    - For the homogenate benchmarking tissue (no core or tissue specimen):
        - [Metadata from GCC] **Tissue aliquot** number will be provided by TPC, e.g., `ST### - #A - ###`
        - [File names at DAC] **Tissue aliquot ID** has the value “**`X`**”, e.g., `ST### - #A - X`
            - To avoid having multiple homogenate aliquots associated with a file name.
    - For the non-homogenate benchmarking tissue (in either core or tissue specimen form):
        - [All] `ST### - #A - ###`
- **Core IDs**: Provided by TPC/NDRI.
    - **Core ID** for the core sample: ID is comprised of a letter between **`A-F`**, followed by a digit between **`1-6`** (ID is associated with the spatial information from within the individual tissue aliquot. See **Figure 1**.)
        - [File names at DAC] For files generated from multiple cores from the same tissue aliquot, **Tissue aliquot** number will be provided by TPC and the **Core ID** will be “MC” e.g., `SMHT### - #A - ###MC`
        - [File names at DAC] For files generated from multiple cores and multiple tissue aliquots, **Tissue aliquot** will be “MA” and the **Core ID** will be “MC” e.g., `SMHT### - #A - MAMC`
    - **Core ID** for the tissue specimen: ID is comprised of a letter between **`S-W`** followed by **`1-9`** (no spatial information associated with the ID).
    - **Core ID** for the homogenate benchmarking tissues and non-solid production tissues: Null value for the ID, i.e., “**`X`**”.
        - [File names at DAC] e.g., `SMHT### - #A - XX` for homogenate benchmarking tissues and `SMHT### - #A - ###X` for non-solid production tissues



Core ID: A letter designation (A-F) to indicate the vertical position followed by an integer to indicate the horizontal position (e.g. A1) within a given tissue aliquot.

Figure 1. Example of spatially numbered tissue aliquots from non-bisected tissue samples. Fixed (pink) or frozen aliquots (green) are numbered separately. “**X**” represents null values to indicate samples that do not get sectioned

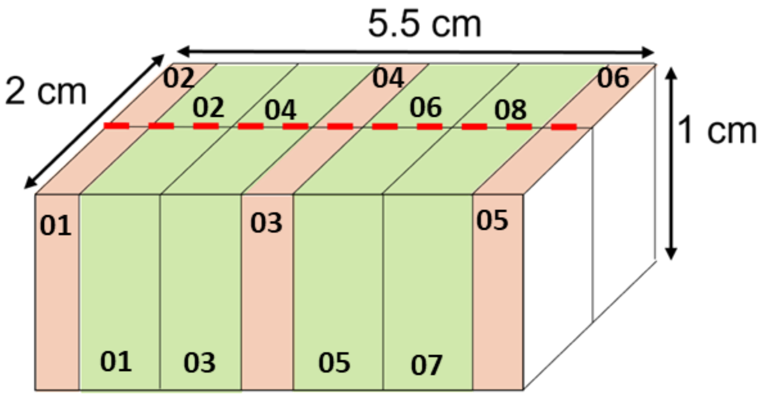


Figure 2. Example of spatially numbered tissue aliquots from longitudinally bisected tissue samples. Fixed (pink) or frozen (green) aliquots are numbered separately, in sequential order starting from the superior end of the sample.

Table 2A. Protocol IDs for SMaHT **benchmarking tissues.**

Sample Protocol ID	**Tissue Name**	**Preservation**	**Notes**
1A	Liver	Snap Frozen	Homogenate and non-homogenate samples

1B	*unassigned*	*N/A*	
1C	Liver	Fixed	
1D	Lung	Snap Frozen	Homogenate and non-homogenate samples
1E	*unassigned*	*N/A*	
1F	Lung	Fixed	
1G	Colon	Snap Frozen	Homogenate and non-homogenate samples
1H	*unassigned*	*N/A*	
1I	Colon	Fixed	
1J	Skin	Snap Frozen	Tissue specimen (~10 cm)
1K	Skin	Snap Frozen	Tissue core from the intact tissue was made (~1 cm)
1L	Skin	Fixed	
1M/N/O/P	*unassigned*	*N/A*	
1Q	Brain, Frontal Lobe	Snap Frozen	Homogenate and non-homogenate samples

Table 2B. Protocol IDs for SMaHT **production tissues**.

Protocol ID	Tissue Name for Container	Preservation
3A	Blood, Whole	Snap Frozen
3B	Buccal Swab	Fresh
3C	Esophagus	Snap Frozen
3D	Esophagus	Fixed
3E	Colon, Ascending	Snap Frozen
3F	Colon, Ascending	Fixed
3G	Colon, Descending	Snap Frozen
3H	Colon, Descending	Fixed
3I	Liver Sample	Snap Frozen
3J	Liver Sample	Fixed

3K	Adrenal Gland, Left	Snap Frozen
3L	Adrenal Gland, Left	Fixed
3M	Adrenal Gland, Right	Snap Frozen
3N	Adrenal Gland, Right	Fixed
3O	Aorta, Abdominal	Snap Frozen
3P	Aorta, Abdominal	Fixed
3Q	Lung	Snap Frozen
3R	Lung	Fixed
3S	Heart, LV	Snap Frozen
3T	Heart, LV	Fixed
3U	Testis, Left	Snap Frozen
3V	Testis, Left	Fixed
3W	Testis, Right	Snap Frozen
3X	Testis, Right	Fixed
3Y	Ovary, Left	Snap Frozen
3Z	Ovary, Left	Fixed
3AA	Ovary, Right	Snap Frozen
3AB	Ovary, Right	Fixed
3AC*	Dermal Fibroblast	Cultured Cells
3AD	Skin, Calf	Snap Frozen
3AE	Skin, Calf	Fixed
3AF	Skin, Abdomen	Snap Frozen
3AG	Skin, Abdomen	Fixed
3AH	Muscle	Snap Frozen
3AI	Muscle	Fixed
3AJ	Brain	Fresh
3AK	Frontal Lobe, Brain, Left hemisphere	Snap Frozen
3AL	Temporal Lobe, Brain, Left hemisphere	Snap Frozen
3AM	Cerebellum, Brain, Left hemisphere	Snap Frozen

3AN	Hippocampus, Brain, Left hemisphere	Snap Frozen
3AO	Hippocampus, Brain, Right hemisphere	Snap Frozen
3AP	Frontal Lobe, Brain, Left hemisphere	Fixed
3AQ	Temporal Lobe, Brain, Left hemisphere	Fixed
3AR	Cerebellum, Brain, Left hemisphere	Fixed
3AS	Frontal Lobe, Brain, Left hemisphere	Fixed
3AT	Hippocampus, Brain, Right hemisphere	Fixed

*3AC = Fibroblasts are isolated from fresh calf skin.

Cell Lines

These cell-line benchmarking samples were prepared by a GCC or TTD, not distributed by TPC. In such cases, the Kit/Sample ID is assigned by DAC as shown in **Table 1**. For cell-line samples that did not follow TPC protocols, the TissueAliquot ID and Core IDs are both “X”:
e.g., `SMHTCOLO829T - X - X`

Sample and Assay Metadata in the File Name

The file name contains the following additional metadata, delimited by a hyphen (“-”) as shown below.

[Project][Kit/Donor ID] **-** [Protocol ID] - [Tissue aliquot & Tissue core IDs] - **[Sex & Age]** **-** **[Sequencing platform & Sequencing assay code]** **-** **[Center code]** - [Unique Accession ID] - [Analysis info] . [File format]

- **Sex & Age**: Sex and age of the donor, “M” (for male) or “F” (for female), followed by the donor’s age in years. “N” and “N” will be used when sex or age is unknown or not applicable (e.g. cell line mixture).
- **Sequencing platform & assay code**: A code consisting of a single alphabetical letter corresponding to a sequencing platform, followed by a 3-digit number designated for a specific experimental assay type (**Tables 3A and 3B**).
- **Center code**: Letters corresponding to the sequencing centers, i.e., GCCs and TTDs that generated the raw sequencing data submitted to DAC (**Table 4**).

Table 3A. Sequencing platform codes.

SMaHT code	**Sequencing platform**
A	Illumina NovaSeq X, Illumina NovaSeq X Plus

B	PacBio Revio HiFi
C	Illumina NovaSeq 6000
D	ONT PromethION 24
E	ONT PromethION 2 Solo
F	ONT MinION Mk1B
G	Illumina HiSeq X
H (Note: Deprecated)	*Illumina NovaSeq X Plus*
I	BGI DNBSEQ-G400
J	Element AVITI
K	Illumina NextSeq 2000
L	PacBio Sequel IIe
M	Ultima Genomics UG 100

Table 3B. Experimental assay codes.

Code	**Assay Name**	**Description**
000		(Null or not-applicable)
[001-100: DNA-based assays]		
001	WGS	DNA, PCR-free, Bulk, Whole genome sequencing (WGS)
002	PCR WGS	DNA, PCR, Bulk, WGS
003	Ultra-Long WGS	DNA, PCR-free, Bulk, Ultra-Long WGS
004	Fiber-seq	DNA, PCR-free, Bulk, Fiber-seq
005	Hi-C	DNA, Bulk, Hi-C
006	Bulk NTSeq	DNA, Bulk, NTSeq
007	CODEC	DNA, Bulk, Duplex-seq, CODEC
008	Bot-seq	DNA, Bulk, Duplex-seq, Bot-seq
009	NanoSeq	DNA, Bulk, Duplex-seq, NanoSeq

010	scNanoSeq	DNA, Single-cell, Duplex-seq, scNanoSeq
011	DLP+	DNA, Single-cell, DLP+
012	Microbulk MALBAC WGS	DNA, Microbulk, MALBAC-amplified WGS
013	Single-cell MALBAC WGS	DNA, Single-cell, MALBAC-amplified WGS
014	Microbulk PTA WGS	DNA, Microbulk, PTA-amplified WGS
015	Single-cell PTA WGS	DNA, Single-cell, PTA-amplified WGS
016	scDip-C	DNA, Single-cell, scDip-C
017	CompDuplex-seq	DNA, Bulk, Duplex-seq, CompDuplex-seq
018	scCompDuplex-seq	DNA, Single-cell, Duplex-seq, scCompDuplex-seq
019	Strand-seq	DNA, Bulk, Strand-seq
020	scStrand-seq	DNA, Single-cell, scStrand-seq
021	HiDEF-seq	DNA, Bulk, Duplex-seq, HiDEF-seq
022	HAT-seq	DNA, Bulk, HAT-seq
023	Microbulk HAT-seq	DNA, Microbulk, PTA-amplified HAT-seq
024	scHAT-seq	DNA, Single-cell, PTA-amplified, HAT-seq
025	VISTA-seq	DNA, Bulk, Duplex-seq, VISTA-seq
026	Microbulk VISTA-seq	DNA, Microbulk, Duplex-seq, VISTA-seq
027	scVISTA-seq	DNA, Single-cell, Duplex-seq, VISTA-seq
028	TEnCATS	DNA, Bulk, TEnCATS
029	L1-ONT	DNA, Bulk, L1-ONT
030	ppmSeq	DNA, Bulk, Duplex-seq, ppmSeq
[101-200: RNA-based assays]		
101	RNA-seq	RNA, Bulk, RNA-seq
102	Kinnex	RNA, Bulk, Kinnex
103	snRNA-seq	RNA, Single-cell, snRNA-seq

104	STORM-seq	RNA, Single-cell, STORM-seq
105	Tranquil-seq	RNA, Single-cell, Tranquil-seq
[201-300: Chromatin-based assays]		
201	ATAC-seq	Chromatin, Bulk, ATAC-seq
202	CUT&Tag	Chromatin, Bulk, CUT&Tag
203	varCUT&Tag	Chromatin, Bulk, varCUT&Tag
204	sc-varCUT&Tag	Chromatin, Single-cell, sc-varCUT&Tag

Table 4. SMaHT data generation center codes.

Code	**Category**	**Institute**	**Contact PI**
bcm	GCC	Baylor College of Medicine	Richard Gibbs
broad	GCC	Broad Institute	Kristin Ardlie
nygc	GCC	New York Genome Center	Soren Germer
uwsc	GCC	University of Washington & Seattle Children’s Hospital	Jimmy Bennett
washu	GCC	Washington University in St. Louis	Ting Wang
bcm1	TTD	Baylor College of Medicine	Chuck Zong
bcm2	TTD	Baylor College of Medicine	Fritz Sedlazeck
bch1	TTD	Boston Children’s Hospital	Christopher Walsh
bch2	TTD	Boston Children’s Hospital	Sangita Choudhury
broad1	TTD	Broad Institute	Fei Chen
cwru	TTD	Case Western Reserve University	Fulai Jin
dfci	TTD	Dana-Farber Cancer Institute	Kathleen Burns
mayo	TTD	Mayo Clinic	Alexej Arbyzov
nyu	TTD	New York University	Gilad Evrony
stfd	TTD	Stanford University	Alexander Urban
umass	TTD	University of Massachusetts	Thomas Fazzio
umich	TTD	University of Michigan	Ryan Mills

uutah	TTD	University of Utah	Gabor Marth
wcnygc	TTD	Weill Cornell Medicine & New York Genome Center	Dan Landau
dac	DAC	Harvard Medical School	Peter Park
tpc	TPC	National Disease Research Interchange (NDRI)	Thomas Bell

File and Other Information in the File Name

The data processing and analysis-related metadata and codes are included in the file name by

[Project][Kit/Donor ID] **-** [Protocol ID] **-** [Tissue aliquot & Tissue core IDs] **-** [Sex & Age] **-** [Sequencing platform & Sequencing assay code] **-** [Center code] **-** **[Unique Accession ID]** **-** **[Analysis info]** **.** **[File format]**

- **Unique Accession ID**: Automatically generated unique accession ID with the "SM" prefix (for SMaHT) for a file stored at DAC. This ID can also be used as an identifier to search the file via the SMaHT data portal.
- **Analysis info and File format**: Information about the analysis used to generate the file can be added here using the following format, delimited by an underscore ("_").
    - For unaligned raw sequence files, a null value, "**X**", is set for the analysis info field.
        - **[Tool name and version; separated by "_"] _ [Genome version] _ [Data type / feature] _ [Other info]**

      e.g. Unaligned FASTQ

      `…–SMAUR192K21D-X.fastq.gz`

      e.g. Unaligned BAM

      `…–SMAUR192K21D-X.bam`

      e.g. BAM aligned against GRCh38 using Sentieon BWA-MEM

      `…–SMAAR1LD95K6-sentieon_bwamem_202308.01_GRCh38.aligned.sorted.bam`

      e.g. VCF of somatic calls using Strelka2 called against the GRCh38 reference genome

      `…–SMAVC1LD95K6-strelka2_4.14_GRCh38.vcf`

        - **Tool name and version (mandatory):** Software/Computational tool name and version used to generate the file, e.g. bwamem_0.7.17.
            - For custom analysis pipelines, tool name is "custom", e.g. custom_1.0
        - **Genome version (optional):** Reference genome version used for aligned sequence files, or variant type for VCF files (**Table 5A**).
        - **Data type or feature (optional; Table 5B)**.
            - Includes additional key info delimited by an underscore ("_").

- For donor-specific assembly files, this can include the haplotype for fasta files and the direction of chain files for reference conversion (SourceToTarget)
- For RNA-Seq data, this can include the GENCODE version and whether counts are for genes or isoforms

Table 5. Genome version (A) and data type (B) tables.

(A)

Reference Genome	**Code**
GRCh38 without ALT contigs	GRCh38
GRCh38 with ALT contigs	GRCh38_ALT
T2T CHM13	CHM13
Donor-specific genome assembly	DSA

(B)

Data Type	**Code**
Reference conversion	[Source]To[Target]
Donor-specific genome assembly haplotype	hapX, hapY, hapX1, hapX2
Gene expression level	gene
Transcript isoform expression level or isoform information	isoform
Junction annotations	junction
Full-length, non-concatemer (FLNC) Kinnex reads	flnc
Aligned consensus Duplex-Seq BAM	consensus

Examples

Benchmarking Cell Line Names

1. COLO829
   `SMHTCOLO829T - X - X - M45`
2. COLO829BL
   `SMHTCOLO829BL - X - X - M45`

3. COLO829 1:50 mixture
   `SMHTCOLO829BLT50 - X - X - M45`
4. HapMap mixture
   `SMHTHAPMAP6 - X - X - NN`
5. iPSC / fibroblast, e.g. LB-IPSC 1
   `SMHTLBIPSC1 - X - X - M29`

File Names

[Note that space was added to make the delimiters (i.e., “-” and “_”) more visible]

[Project][Kit/Donor ID] **-** [Protocol ID] **-** [Tissue aliquot & Tissue core IDs] **-** [Sex & Age] **-** [Sequencing platform & Sequencing assay code] **-** [Center code] **-** [Unique Accession ID] **-** [Analysis info] **.** [File format]

Files Generated from Benchmarking Samples

Bulk WGS on Illumina NovaSeqX - FASTQ file of the COLO829T-COLO829BL 1:50 admixture from BCM-GCC:
`SMHTCOLO829BLT50 - X - X - M45 - A001 - bcm - SMAUR123JK01 - X . fastq.gz`

FiberSeq on PacBio HiFi - unaligned BAM file of the COLO829 tumor cell line from UWSC-GCC:
`SMHTCOLO829T - X - X - M45 - B003 - uwsc - SMAUR89QQ3LA - X . bam`

FiberSeq on PacBio HiFi - aligned BAM file of the COLO829 tumor cell line from UWSC-GCC, aligned to GRCh38, sorted by read name, and phased:
`SMHTCOLO829T - X - X - M45 - B003 - uwsc - SMAFI57S7SLA - pbmm_2.6_GRCh38 . aligned.sorted.phased.bam`

Bulk WGS on Ultima Genomics - SNVs called using Strelka2 from COLO829 1:50 admixture against COLO829BL from NYGC-GCC:
`SMHTCOLO829BLT50 - X - X - M45 - M001 - nygc - SMAVC29A1QR7 - strelka_2.9_GRCh38 . vcf`

Bulk WGS on Illumina NovaSeqX - FASTQ file of the HapMap mixture from Broad-GCC:
`SMHTHAPMAP6 - X - X - NN - A001 - broad - SMAUR89XAY38 - X. fastq.gz`

- **Aligned CRAM file from the FASTQ file above:**
  `SMHTHAPMAP6 - X - X - NN - A001 - broad - SMAFI33A9K6R - sentieon_bwamem_202308.01_GRCh38 . aligned.sorted.cram`
- **VCF that contains HapMap mixture-specific variants identified by running Mutect 2 on the BAM file above:**
  `SMHTHAPMAP6 - X - X - NN - A001 - broad - SMAVC3R123U9 - mutect_4.1.4.1_GRCh38 . vcf`

DSA - COLO829 1:50 admixture against COLO829BL from haplotype with X chromosome:
`SMHTCOLO829BLT50 - X - X - M45 - F001 - nygc - SMASF29A1QR7 - hifiasm_2.4_hapX . fasta`

Accompanying chain file for DSA COLO829 1:50 admixture against COLO829BL:
`SMHTCOLO829BLT50 - X - X - M45 - F001 - nygc - SMASF29A1QR7 - softwareVersion_GRCh37ToDSA . chain.gz`

Bulk WGS on Illumina NovaSeq X - CRAM file of the benchmarking tissue from UWSC-GCC:
`ST002 - 1G - XX - M74 - A001 - uwsc - SMAFIQUWOL9M - sentieon_bwamem_202308.01_GRCh38 . aligned.sorted.cram`

Files Generated from Production Samples

Bulk WGS on Illumina NovaSeqX Plus - CRAM file of a single production tissue core from UWSC-GCC:
`SMHT001 - 3E - 001A1 - M42 - A001 - uwsc - SMAFIQUWOL9M - sentieon_bwamem_202308.01_GRCh38 . aligned.sorted.cram`

Bulk WGS on Illumina NovaSeqX Plus - CRAM file of multiple production tissue cores from the same tissue aliquot from UWSC-GCC:
`SMHT001 - 3E - 001MC - M42 - A001 - uwsc - SMAFIQUWOL9M - sentieon_bwamem_202308.01_GRCh38 . aligned.sorted.cram`

Bulk WGS on Illumina NovaSeqX Plus - CRAM file of multiple production tissue cores from multiple tissue aliquots from UWSC-GCC:
`SMHT001 - 3E - MAMC - M42 - A001 - uwsc - SMAFIQUWOL9M - sentieon_bwamem_202308.01_GRCh38 . aligned.sorted.cram`

Bulk RNA-Seq on Illumina Novaseq 6000 of a single production tissue core from BCM-GCC:

- **BAM file:**
  `SMHT001 - 3E - 001A1 - M42 - C101 - bcm - SMAFIQUWOL9M - star_2.7.10b_GRCh38_gencode_v47 . aligned.sorted.bam`

- **TSV file of gene counts**
  `SMHT001 - 3E - 001A1 - M42 - C101 - bcm - SMAFIQUWOL9M - rsem_v1.3.3_GRCh38_gencode_v47_gene . tsv`
- **TSV file of isoform counts**
  `SMHT001 - 3E - 001A1 - M42 - C101 - bcm - SMAFIQUWOL9M - rsem_v1.3.3_GRCh38_gencode_v47_isoform . tsv`
- **Tar file of RNASeQC output**
  `SMHT001 - 3E - 001A1 - M42 - C101 - bcm - SMAFIQUWOL9M - rnaseqc_v2.4.2_GRCh38_gencode_v47 . tar.gz`