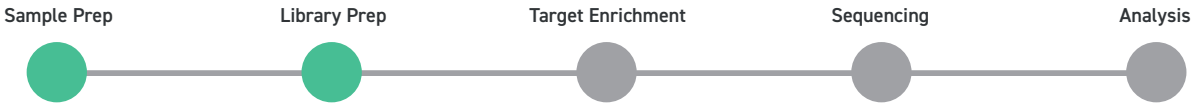
Twist Human Sample ID Protocol with EF 2.0 Library Preparation Kit and Twist Universal Adapter System

For use with the Twist NGS Workflow

This Twist Human Sample ID Library Preparation Kit provides the reagents needed to prepare sample-tracking libraries from either fresh, whole blood or from purified genomic DNA (gDNA) using multiplex PCR amplification and ligation of adapters from the Twist Universal Adapter system. The Twist Universal Adapter system consists of Twist Universal Adapters and the Twist Unique Dual Indexed (UDI) Primers. This user protocol details the steps for generating the amplified, indexed libraries needed for sequencing on Illumina next-generation sequencing (NGS) systems. This Sample ID library preparation protocol is optimized for use in combination with libraries prepared with Twist Enzymatic Fragmentation Library Preparation protocols using Twist UDI indexes and enriched with the Twist Exome 2.0 panels and Twist Target Enrichment Kits. Target enrichment with an alternate panel will require the addition of probes that include the Sample ID targets. This protocol should only be performed with reagents specified or their equivalents.



For Research Use Only. *Not intended for use in diagnostic procedures.*

PROTOCOL COMPONENTS

CATALOG #	NAME	DESCRIPTION	STORAGE
105895, 105896, 105897, 105898: 96 rxn	Twist Sample ID Kit with EF 2.0 and UDI Plate A, B, C, or D	Reagents for Sample ID Library Preparation	—
	Twist Human Sample ID Kit	• 4x Sample ID Master Mix • Sample ID Primer Mix	–20°C
	Twist Library Preparation EF Kit 1, 2.0	• Frag/AT Enzymes • Frag/AT Buffer • Ligation Master Mix • Equinox Library Amp Mix (2x) • P5/P7 Primers (10x)	–20°C
	Twist Library Preparation Kit 2	DNA Purification Beads	2–8°C
	Twist Universal Adapter System – TruSeq Compatible	Twist Universal Adapters and Twist UDI Primers, provides unique dual-indexed combinations with 1 reaction per index pair	–20°C

LEGAL

This document may contain references to other third-party resources such as sources of information, hardware or software, products, or services and/or web sites owned or licensed by third parties. Twist Bioscience does not control or take responsibility for any third-party resources, including, without limitation, the accuracy, reliability, copyright compliance, compatibility, performance, legality, or any other aspect of third-party resources. The inclusion of such resources in this document does not imply endorsement by Twist Bioscience of any third-party resources.

Certain processes described in this document may be subject to patent rights or licenses in local jurisdictions, including those owned or licensed by parties other than Twist Bioscience. Purchase of this product does not include a license to perform any such processes. Users of this product may, therefore, be required to obtain a patent license depending upon the particular application and country in which the product is used before performing such processes.

Twist Bioscience, the various Twist logos, and the Twist NGS logo used herein are trademarks of Twist Bioscience Corporation. All other trademarks are the property of their respective owners.

INTENDED USE

This product is for research use only. This product is not intended for the diagnosis, prevention, or treatment of a disease or condition. Twist Bioscience assumes no liability regarding use of the product for applications in which it is not intended.

TABLE OF CONTENTS

MATERIALS SUPPLIED BY USER

PRODUCT	SUGGESTED SUPPLIER
REAGENTS AND CONSUMABLES	
Ethanol (200 proof)	—
Molecular biology grade water	—
10 mM Tris-HCl pH 8	—
Buffer EB (optional)	Qiagen
1.5-ml microcentrifuge tubes	VWR
Thin-walled PCR 0.2 mL strip tubes	Eppendorf
96-well thermal cycling plates	VWR
96-well compatible magnetic plate	Alpaqua, Permagen Labware
Qubit dsDNA High Sensitivity Quantitation Assay	Thermo Fisher Scientific
Qubit dsDNA Broad Range Quantitation Assay (optional)	Thermo Fisher Scientific
Agilent DNA 1000 Kit	Agilent Technologies
EQUIPMENT	
Pipettes and tips	—
Vortex Mixer	—
Benchtop mini-centrifuge for 0.2 mL tubes	—
Thermal cycler (96-well) with heated lid	—
Fluorometer (Qubit 3.0)	Thermo Fisher Scientific
2100 Bioanalyzer or equivalent	Agilent Technologies

GENERAL NOTES AND PRECAUTIONS

Wear appropriate protective equipment (lab coat, gloves, and protective glasses or goggles) at all times when performing this protocol.

For best results, read this document before performing the protocol, and follow the instructions provided. Twist cannot guarantee the performance of the Twist Human Sample ID Library Preparation Kit and the Twist Universal Adapter System if modifications are made to the protocol.

This library preparation method may yield more material than needed for sequencing. Excess product can be stored at –20°C for later use.

Test the compatibility of your thermal cycler and PCR tubes by incubating at 95°C for up to 5 minutes to ensure the PCR tubes do not crack under heat and pressure. Adjust the tightness of the thermal cycler lid and/or use a spacer specific to the thermal cycler model.

GUIDELINES FOR BLOOD SAMPLES

NOTE: These guidelines are for users setting up blood work for the first time.

- Users handling blood should undergo BSL-2 level, or higher, training.
- Handling of samples prior to the completion of the multiplex PCR should be performed in a BSL-2 compliant biosafety cabinet or area.
- Whole blood samples from EDTA-preserved and Streck cfDNA blood collection tubes have been validated. Further optimization may be required if using alternate preservatives or collection methods.
- Thoroughly homogenize the blood sample by inverting the tube.
- The recommended input is 2 µl of whole blood.
  - Varying the volume and/or blood source will require additional optimization.

GUIDELINES FOR PURIFIED gDNA SAMPLES

- Use the Thermo Fisher Scientific Qubit Quantitation Assay to accurately quantify input purified gDNA.
  - Either the Qubit High Sensitivity Assay or Qubit Broad Range Assay can be used depending on the gDNA extraction protocol employed and the typical yields experienced by the user.
- Measuring DNA concentration by absorbance at 260 nm is not recommended.
- Input DNA should be suspended in Molecular Biology Grade Water, 10 mM Tris-HCl pH 8.0, or Buffer EB.
- The recommended input is 10 ng of high quality gDNA
  - Lower quality gDNA may require further optimization of the Sample ID protocol
- For technical support, contact customersupport@twistbioscience.com.

PROTOCOL OVERVIEW

This protocol begins with whole blood or purified genomic DNA (gDNA) and generates amplified, indexed libraries for sequencing. It features DNA-end preparation and Twist Universal Adapters with UDI primers. This protocol allows the user to perform Sample ID library preparation (Steps 1–4) in 5 hours.

	SAMPLE ID LIBRARY PREPARATION WITH UNIVERSAL ADAPTERS AND UDI PRIMERS (2 µL WHOLE BLOOD OR 10 NG GENOMIC DNA STARTING MATERIAL)	**TIME**
STEP 1	**PCR Amplified Sample ID Targets** Amplified Inserts	**2 hours**
STEP 2	**Perform End Repair and dA-tailing** dA-tailed DNA Amplicons	**1 hour**
STEP 3	**Ligate Twist Universal Adapters and Purify** Amplicon libraries ready for indexing	**1 hour**
STEP 4	**PCR Amplified Using Twist UDI Primers, Purify, and Perform QC** Amplified indexed libraries	**1 hour**

STEP 1 PCR AMPLIFIED SAMPLE ID TARGETS

Perform multiplex PCR from whole blood or purified gDNA to enrich for Sample ID targets.

Reagents Required

- DNA Input
  - Whole blood: 2 µl per sample
    or
  - Purified gDNA: 10 ng per sample
- Molecular biology grade water
- From the Twist Human Sample ID kit
  - Sample ID Primer Mix
  - 4x Sample ID Master Mix

Before You Begin

- If possible, transfer a 10–50 µl aliquot of whole blood into a PCR tube or well of a 96-well plate to facilitate accurate transfer into the PCR reaction.
  - Leave whole blood aliquots at room temperature or on a cold block consistent with blood collection tube storage conditions
- Thaw and place on ice or cold block
  - gDNA (if using as DNA input)
  - Sample ID Primer Mix
  - 4x Sample ID Master Mix

PREPARE THE THERMAL CYCLER, SAMPLES AND REAGENTS

1.1

Program the thermal cycler with the following conditions. Set the temperature of the heated lid to 105°C. Start the program to pre-heat the thermal cycler.

	STEP	TEMPERATURE	TIME	NUMBER CYCLES
1	(HOLD)	98°C	HOLD	1
2	Initialization	98°C	3 minutes	1
3	Denaturation	98°C	15 seconds	23
	Annealing	65°C	3 minutes	
	Extension	72°C	30 seconds	
4	Final Extension	72°C	1 minute	1
5	Final Hold	4°C	HOLD	1

PERFORM THE SAMPLE ID PCR

1.2 Invert 4x Sample ID Master Mix≥5 times before use. Do not vortex. Pulse-spin to collect all liquid at the bottom of the tube.

1.3 Vortex the Sample ID Primer mix for 5 seconds. Pulse-spin to collect all liquid at the bottom of the tube.

1.4 Add 6.25 µl of the 4x Sample ID Master Mix to each well or 0.2 ml, thin-walled PCR tube.
NOTE: Assemble the PCR reaction(s) on ice or cold block.

1.5 Add 1 µl of the Sample ID Primer Mix to each well or tube from Step 1.3.

1.6 Add either 2 µl of whole blood or 10 ng of purified gDNA to each well or tube from Step 1.5.
NOTE: Ensure the whole blood is homogeneously resuspended.

1.7 Add molecular grade water up to 25 µl for each well or tube from Step 1.6 and mix well by gentle pipetting.

1.8 Cap the tube(s) or seal the plate and pulse-spin the samples. Then immediately transfer to the thermal cycler.

1.9 Proceed to the Initialization step of the PCR program.
NOTE: When there are 30 minutes remaining on the thermal cycler program, prepare reagents for Step 2: Perform End Repair and dA-Tailing (see Before You Begin).

1.10 When the thermal cycler program is complete and the sample block has reached 4°C, remove the samples from the block and place them on ice (see Step 2.3).

REAGENT	WHOLE BLOOD	PURIFIED gDNA
4x Sample ID Master Mix	6.25 µl	6.25 µl
Sample ID Primer Mix	1 µl	1 µl
Whole Blood or Purified gDNA (10 ng)	2 µl —	— up to 17.75 µl
Molecular Grade Water	15.75 µl	up to 25 µl
Total Volume	25 µl	25 µl

Guide for assembling PCR reactions

STEP 2 PERFORM END REPAIR AND dA-TAILING

Perform End Repair and dA-tailing to generate dA-tailed amplicons.

Reagents Required

- DNA Input: 1 µl per sample
- 10 mM Tris-HCl, pH 8 or Buffer EB (cold)
- From the Twist Library Preparation EF Kit 1, 2.0
  - Frag/AT Enzyme Mix
  - Frag/AT Buffer

Before You Begin

- Thaw and place on ice or cold block
  - 10 mM Tris-HCl or Buffer EB
  - PCR product from Step 1
  - Frag/AT Buffer
  - Frag/AT Enzyme Mix

PREPARE THE THERMAL CYCLER AND REAGENTS

2.1 Program the thermal cycler with the following conditions. Set the temperature of the heated lid to 105°C. Start the program to pre-heat the thermal cycler.

STEP	TEMPERATURE	TIME
1	65°C	HOLD
2	65°C	30 minutes
3	4°C	HOLD

2.2 Pre-chill 39 µl of 10 mM Tris-HCl in a well or 0.2 ml, thin-walled PCR tube per sample.

2.3 Place PCR reaction from Step 1 on ice or cold block for 15 seconds.

2.4 Pulse-spin the tube(s) for >15 seconds to collect cellular debris at the bottom of the tube.
NOTE: If using purified gDNA, Steps 2.3–2.4 can be skipped.

2.5 Pipette 1 µl of the cleared PCR reaction from Step 2.4 into the 39 µl pre-chilled Tris-HCl buffer from Step 2.2. Thoroughly mix by gentle pipetting (x3) on ice. Keep on ice until after Step 2.8.

PERFORM END REPAIR AND dA-TAILING

2.6 Vortex the Frag/AT Buffer for 5 seconds. Pulse-spin to collect all liquid at the bottom of the tube.

2.7 Invert the Frag/AT Enzyme Mix a minimum of 10 times to homogenize or briefly vortex to ensure complete mixing. Pulse-spin to collect all liquid at the bottom of the tube.

2.8 Prepare an End Prep mix in a 1.5 ml microfuge tube on ice. Use the volumes listed below. Homogenize the mastermix by moderately vortexing for 5 seconds or by pipetting the mastermix up and down 10 times (avoid formation of bubbles).

REAGENT	VOLUME PER REACTION*
Frag/AT Buffer	4 µl
Frag/AT Enzyme Mix	6 µl
Total	**10 µl**

**Prepare a master mix for multiple reactions.*

2.9 Add 10 µl of the End Prep mix (from Step 2.8) to each 40 µl diluted multiplex PCR sample tube or well. On ice, pipette 25 µl of each reaction mixture up and down 10 times (avoid formation of bubbles). Cap the tube(s) or seal the plate and keep the reaction on ice.
NOTE: Avoid removing samples from ice as much as possible to maintain desired amplicon sizes.

2.10 Pulse-spin the tube(s) or plate and immediately transfer to the pre-heated thermal cycler.
NOTE: If the thermal cycler is not immediately adjacent to the centrifuge in Step 2.10, immediately place tube(s) or plate on ice during transport.

2.11 Proceed to the 30-minute, 65°C incubation step of the thermal cycler program (refer to the table in Step 2.1 above).
NOTE: While the thermal cycler program is running, prepare the reagents for Step 3: Ligate Twist Universal Adapters and Purify (see Before you Begin).

2.12 When the thermal cycler program is complete and the sample block has reached 4°C, remove the samples from the block and place them on ice.

PROCEED IMMEDIATELY TO STEP 3: LIGATE TWIST UNIVERSAL ADAPTERS AND PURIFY

STEP 3 LIGATE TWIST UNIVERSAL ADAPTERS AND PURIFY

Ligate Twist Universal Adapters to the dA-tailed DNA fragments from Step 2 and purify to generate Sample ID libraries ready for index introduction through amplification in Step 4.
NOTE: To properly demultiplex Sample ID libraries from target enriched libraries after sequencing, use separate Twist UDI indexes. For example, if Twist UDI Plate A is used for 96 target enriched libraries, use Twist UDI Plate B for the cognate Sample ID libraries.

Reagents Required

- dA-tailed amplicons (from Step 2.12)
- Ethanol
- 10 mM Tris-HCl, pH 8 or Buffer EB
- From the Twist Library Preparation EF Kit 1, 2.0:
  - Ligation Master Mix
- From the Twist Universal Adapter System:
  - Twist Universal Adapters
- From the Twist Library Preparation Kit 2:
  - DNA Purification Beads

Before You Begin

- Thaw and place on ice or cold block
  - Twist Universal Adapters (tube; utilized for all samples)
  - Ligation Master Mix
- Prepare 1 ml 80% Ethanol for each sample (for use in both Steps 3 and 4 of the protocol)
- Equilibrate DNA Purification Beads to room temperature for ≥30 minutes (for use in both Steps 3 and 4 of the protocol)
- Program a thermal cycler to incubate the samples at 20°C with the heated lid set to minimum temperature or turned off. Start the program so that the cycler is at 20°C when the samples are prepared.

LIGATE TWIST UNIVERSAL ADAPTERS

3.1 Add 5 µl Twist Universal Adapters into each sample well or tube containing the dA-tailed DNA fragments from Step 2. Mix gently by pipetting and keep on ice.

3.2 Invert the Ligation Master Mix a minimum of 10 times until homogenized and place on ice.
NOTE: Do not vortex the Ligation Master Mix.

3.3 Add 20 µl of Ligation Master Mix to each sample from Step 3.1. Pipette a minimum of 40 µl up and down 10 times to ensure complete mixing. Seal or cap the sample plate or tube(s) and pulse-spin to ensure all solution is at the bottom of the tube.

3.4 Incubate the ligation reaction at 20°C for 15 minutes in the thermal cycler, then move the samples to the bench top. Proceed to the Purify step.

⚠ IMPORTANT: Turn off the heated lid or set to minimum.

NOTE: While the thermal cycler program is running, prepare the reagents for Step 4: PCR Amplify Using the Twist UDI Primers, Purify, and Perform QC.

PURIFY

3.5 Vortex the pre-equilibrated room temperature DNA Purification Beads until well mixed.

3.6 Add 60 µl of homogenized (0.8x) DNA Purification Beads to each ligation sample from Step 3.4. Mix well by vortexing.

3.7 Incubate the samples for 5 minutes at room temperature.

3.8 Place the samples on a magnetic plate for 1 minute or until the supernatant is clear.

3.9 The DNA Purification Beads form a pellet, leaving a clear supernatant. Without removing plate or tube(s) from the magnetic plate, remove and discard the supernatant.

3.10 Wash the bead pellet by gently adding 200 µl freshly prepared 80% ethanol (do not disturb the pellet). Incubate for 1 minute, then remove and discard the ethanol.

3.11 Repeat Step 3.10, for a total of two washes, while keeping the sample(s) on the magnet.

3.12 Carefully remove all remaining ethanol with a 10-µl pipette, making sure not to disturb the bead pellet.

3.13 Air-dry the bead pellet on the magnetic plate for 5 minutes or until the bead pellet is dry. Do not overdry the bead pellet.

3.14 Remove the plate or tube(s) from the magnetic plate and add 17 µl, 10 mM Tris-HCL or Buffer EB to each sample. Vortex to mix, making sure to resuspend the entire bead pellet.

3.15 Incubate at room temperature for 2 minutes. Then pulse-spin the sample plate or tube(s).

3.16 Place the plate or tubes on a magnetic plate and let stand for 3 minutes or until the beads form a pellet.

3.17 Transfer 15 µl of the clear supernatant containing the ligated libraries to a clean thin- walled PCR 0.2-ml strip-tube or well of a 96-well thermal cycling plate, making sure not to disturb the bead pellet.

PROCEED TO STEP 4: PCR AMPLIFIED USING TWIST UDI PRIMERS, PURIFY, AND PERFORM QC

STEP 4 PCR AMPLIFIED USING THE TWIST UDI PRIMERS, PURIFY, AND PERFORM QC

Amplify the adapted Sample ID libraries with Twist UDI Primers, purify them, and perform quality control (QC) analysis to complete the protocol.

Reagents Required

- Ligated libraries (from Step 3.17)
- 80% Ethanol (from Step 3)
- Equilibrated DNA Purification Beads (from Step 3)
- 10 mM Tris-HCl, pH 8 or Buffer EB
- From the Twist Library Preparation EF Kit 1, 2.0:
  - Equinox Library Amp Mix (2x)
- From the Twist Universal Adapter System:
  - Twist UDI Primers

⚠ IMPORTANT: Use of P5/P7 Primers (10x) tubes PNs 104127, 103842 contained in the Twist Library Preparation EF Kit 1, 2.0 is not required. Using these primers with the Twist Universal Adapter System will result in a failed PCR amplification.

Before You Begin

- Thaw and place on ice or cold block
  - Equinox Library Amp Mix (2x)
  - Twist UDI Primers (plate with single-use primers)

PREPARE THE THERMAL CYCLER

4.1

Program a thermal cycler with the following conditions. Set the temperature of the heated lid to 105°C.

	STEP	TEMPERATURE	TIME	# OF CYCLES
1	Initialization	98˚C	45 seconds	1
2	Denaturation	98˚C	15 seconds	8*
	Annealing	60˚C	30 seconds	
	Extension	72˚C	30 seconds	
3	Final Extension	72˚C	1 minute	1
4	Final Hold	4˚C	HOLD	–

**8 cycles is recommended when starting with 2 µl whole blood or 10 ng high quality purified gDNA. This may need to be optimized for other input types, volumes and sample quality.*

4.2

Add 10 µl of Twist UDI Primer from the provided 96-well plate to each of the Sample ID libraries from Step 3.17 and mix well by gentle pipetting.

4.3 Add 25 µl of Equinox Library Amp Mix (2x) to the Sample ID libraries from Step 4.2 and mix well by gentle pipetting.
NOTE: Invert Equinox Library Amp Mix (2x) 5 times before use. Do not vortex.

4.4 Pulse-spin sample plate or tube(s) and immediately transfer to the thermal cycler. Start the program.

4.5 Remove the sample(s) from the block when the thermal cycler program is complete. Proceed to purification.

PURIFY

4.6 Vortex the pre-equilibrated DNA Purification Beads until mixed.

4.7 Add 50 µl (1x) of homogenized DNA Purification Beads to each ligation sample from Step 4.5. Mix well by vortexing.

4.8 Incubate the samples for 5 minutes at room temperature.

4.9 Place the samples on a magnetic plate for 1 minute. The DNA Purification Beads form a pellet, leaving a clear supernatant.

4.10 Without removing plate or tubes from the magnetic plate, remove and discard the supernatant.

4.11 Wash the bead pellet by gently adding 200 µl freshly prepared 80% ethanol (do not disturb the pellet), incubate for 1 minute, then remove and discard the ethanol.

4.12 Repeat Step 4.11, for a total of 2 washes, while keeping the samples on the magnet.

4.13 Carefully remove all remaining ethanol with a 10 µl pipette, making sure not to disturb the bead pellet.

4.14 Air-dry the bead pellet on the magnetic plate for 5 minutes or until the bead pellet is dry. Do not overdry the bead pellet.

4.15 Remove the plate or tubes from the magnetic plate and add 22 µl water, 10 mM Tris-HCl pH 8, or Buffer EB to each sample. Mix by pipetting until homogenized.

4.16 Incubate at room temperature for 2 minutes.

4.17 Place the plate or tubes on a magnetic plate and let stand for 3 minutes or until the beads form a pellet.

4.18 Transfer 20 µl of the clear supernatant containing the Amplified Indexed Libraries to a clean thin-walled PCR 0.2-ml strip-tube or well of a 96-well thermal cycling plate, making sure not to disturb the bead pellet.

PERFORM QC

4.19

Quantify and validate the size range of each library using the Thermo Fisher Scientific Qubit dsDNA High Sensitivity Quantitation Assay and Agilent DNA 1000 Assay.

Using 2 µl of fresh, whole blood or 10 ng of purified gDNA amplified with 8 cycles in the indexing PCR should result in a final concentration of ≥ 2ng/µl. Concentrations below 2 ng/µl may be caused by low white blood cell count in a sample or may reflect inefficient sample preparation. Under standard conditions, several peaks should be observed with a mode size of ~430 bp using a range setting of 150–1,000 bp. Library concentrations <2 ng/µl may require the Agilent DNA High Sensitivity Assay to confirm the size profile.

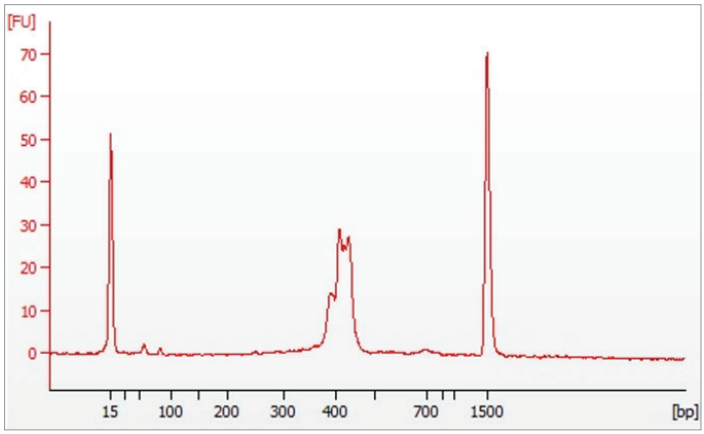


Figure 1. Representative electropherogram of a purified Sample ID library generated with 2 µl of whole blood.

APPENDIX | SECTION 1: SAMPLE ID CONTENT

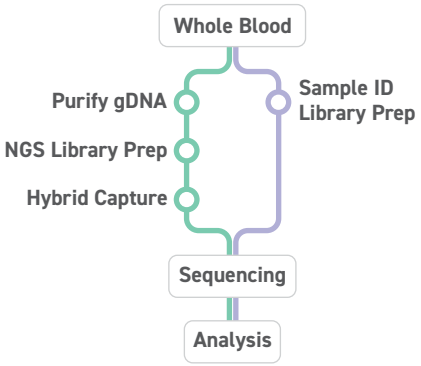
Table 1. Sample ID SNP loci and MAF

CHROMOSOME (HG38)	START	END	SNP_ID	MAF
chr1	78112492	78112493	rs6666954	0.325851
chr10	98459556	98459557	rs10883099	0.506933
chr12	884763	884764	rs7300444	0.420008
chr12	67801314	67801315	rs10748087	0.552429
chr13	38859468	38859469	rs9532292	0.373783
chr15	34236746	34236747	rs4577050	0.653517
chr16	70269676	70269677	rs2070203	0.470352
chr17	73201608	73201609	rs1037256	0.521692
chr18	23833904	23833905	rs9962023	0.658658
chr19	10156400	10156401	rs2228611	0.518105
chr2	146839404	146839405	rs4411641	0.457752
chr2	168932505	168932506	rs497692	0.544797
chr2	227032259	227032260	rs10203363	0.472765
chr20	6119440	6119441	rs10373	0.517533
chr20	49684608	49684609	rs6512586	0.511819
chr21	42903479	42903480	rs4148973	0.558026
chr22	20787011	20787012	rs4675	0.46092
chr3	4362082	4362083	rs2819561	0.590071
chr3	60912761	60912762	rs11130795	0.437418
chr4	185078388	185078389	rs6841061	0.592314
chr6	130827722	130827723	rs9388856	0.378067
chr6	146434003	146434004	rs2942	0.502997
chr7	48410559	48410560	rs17548783	0.480745
chr8	93923708	93923709	rs4735258	0.470757
chr8	106224051	106224052	rs1393978	0.556155
chr9	87447907	87447908	rs12682834	0.643669
chr9	97428497	97428498	rs1381532	0.51109
chrY	13479656	13479657	rs2032653	0.079689
chrY	20759881	20759882	rs34581739	0.072943
chrX	11296917	11296918	AMELX	NaN
chrY	6869907	6869908	AMELY	NaN

APPENDIX | SECTION 2: SAMPLE ID POOLING GUIDELINES

The Sample ID library is designed for pooling with target enriched libraries containing the loci listed in Table 1. Recommendations for pooling may depend on factors including, but not limited to, sample quality, hybrid capture panel size, capture efficiency, and sequencer loading and sequencing quality.

It is recommended to target ~93,000 paired-end (PE) reads on a flow cell to accommodate variability in library prep and sequencer loading conditions.



METHOD 1: ESTIMATE SAMPLE ID LIBRARIES LOADING BASED ON EXPECTED TOTAL READS

Example:

- NextSeq 550 High Output kit, 300 cycle
- High Output Flow Cell
- Up to 800 M Paired-End reads*

EXPECTED TOTAL PE READS	DILUTION FOR ~93,000 PE READS
300 Million	~1/3200
500 Million	~1/5300
800 Million	~1/8600

**https://www.illumina.com/systems/sequencing-platforms/nextseq/specifications.html*

1. Dilute Target Enriched Samples.

a. Dilute pooled target enriched libraries to 4.4 nM in 9 µl.

2. For each Sample ID library, dilute to:

- 12.5 pM if the NextSeq 550 output is expected to yield 300 M PE reads, or
- 7.55 pM if the NextSeq 550 output is expected to yield 500 M PE reads, or
- 4.65 pM if the NextSeq 550 output is expected to yield 800 M PE reads, or adjust accordingly for sequencing yields typical for the user

3. Pool the Sample ID libraries in >1 µl such that the desired concentrations in step 2 are maintained.

4. Combine 1 µl of the pooled Sample ID libraries with 9 µl of the pooled target enriched libraries.

5. Treat this final pool as a 4 nM concentration and proceed with sample loading preparation for the sequencer as normal.

METHOD 2: COMBINE AS A RATIO TO THE EXOME 2.0 HYBRID CAPTURE LIBRARY

Example: Combine 12 Exome 2.0 hybrid capture libraries with 12 Sample ID libraries.

EXOME PARAMETER	EXOME VALUE
Panel Size	~36.5 Mb
Targeted Raw Coverage	200X
Total Bases Needed	7.3 Gb
Read Length	2 x 100 bp
Total Reads Needed	73 M, PE reads
SAMPLE ID PARAMETER	**SAMPLE ID VALUE**
Total Reads needed	93 k, PE reads
POOLING	
Ratio (Exome:Sample ID)	784:1

1. Dilute libraries to 4.3 nM. (After the final pooling, the final concentration will be close to 4 nM in this example)

2. Pool the 12 exome libraries together

3. Pool the 12 Sample ID libraries together in a separate tube

a. Dilute the pooled Sample ID libraries 1/60 with molecular grade water

4. Combine the pooled exomes and the pooled, diluted Sample ID libraries at a ratio of 12:1

a.The final ratio for any give Sample ID to its paired exome library will be [(1/60)x(1/13)] = 1/780, or 1:781.

APPENDIX | SECTION 3: BFX VARIANT CALLING GUIDELINES

COMPARING AMPLICON AND HYBRID-CAPTURE SAMPLES

This guide provides step-by-step instructions for analyzing FASTQ files produced from a next-generation sequencing run of the SampleID kit, including instructions for comparing variant calls from the sample to existing variant calls produced from a separate sample.

STARTING POINT

You will need the following software to proceed with this analysis*

- seqtk for downsampling reads (https://github.com/lh3/seqtk) (optional)
- cutadapt for trimming adapter sequences (https://github.com/marcelm/cutadapt)
- bwa for alignment (https://github.com/lh3/bwa), or whatever aligner is used for primary analysis on capture samples.
- samtools for processing alignments (https://github.com/samtools/)
- GATK4 for assessing quality metrics from the sequencing, and for performing variant calling (https://github.com/broadinstitute/gatk)

In addition to the software listed above, you will need:

- A copy of the human reference genome, including any indexes necessary for alignment
- FASTQ files from a run of the SampleID kit for analysis
- Variant calls (as a VCF) from a matched hybrid capture sample
- BED files for the amplicon and target sites present in the SampleID kit (available from Twist's website)

**Twist does not assist with configuring, compiling, executing, or troubleshooting third-party software packages. Anyone seeking such assistance will need to obtain help from their own IT professionals. Twist does not guarantee the output of any third-party software package is accurate, nor does Twist warranty that any third-party software package will be fit for your particular use.*

DOWNSAMPLE ALIGNMENTS (SEQTK)

To keep analyses reproducible, and to minimize the computational time needed for subsequent steps, we recommend downsampling reads to a fixed number of reads. For this purpose, we recommend `seqtk sample`. We have generally performed analysis against a depth of 1000x over each of the 31 target sites, for a total of 31,000 reads. An example pair of commands is shown below to sample to this depth.

```
seqtk sample -2 -s 42 input_R1.fastq 31000 > sampled_R1.fastq
seqtk sample -2 -s 42 input_R2.fastq 31000 > sampled_R2.fastq
```

Where 42 is a random seed (must be identical for both files), -2 specifies two-pass mode (more accurate but slower), and 31,000 is the total number of required reads.

TRIM FASTQ FILES (CUTADAPT)

First, FASTQ files should be trimmed to remove adapter sequences. We recommend cutadapt for this purpose. This is useful both for improving alignment quality, and for eliminating any adapter-dimer containing reads from the input pool

```
cutadapt -A AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT -a
AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC -o trimmed_R1.fastq.gz -p
trimmed_R2.fastq.gz {input_R1.fastq.gz} {input_R2.fastq.gz} -m {read_length - 40}
```

Where read_length is the input read length, and input_R1.fastq.gz and input_R2.fastq.gz are the R1 and R2 input files respectively (possibly after sampling). We have found that discarding reads that lose more than 40 base pairs does a good job of distinguishing between adapter dimers and full-length reads.

ALIGN TO REFERENCE GENOME, AND SORT ALIGNMENTS (BWA/SAMTOOLS)

Using the trimmed FASTQ files, we will now align the files to the human reference genome. We recommend using the same aligner and reference genome that is used for the paired capture samples. For example, if BWA was used to align the reads to the reference for the paired capture samples, the following command could be used to align to the reference

```
bwa mem {path_to_genome_index} trimmed_R1.fastq.gz
trimmed_R2.fastq.gz | samtools view -bh | samtools sort > sorted_alignments.bam
```

Where path_to_genome_index is a path to the BWA genome index on the local machine, and the R1 and R2 files are the trimmed output from the previous step.

GENERATE TARGETED SEQUENCING METRICS (GATK4/COLLECTTARGETEDPCRMETRICS)

We highly recommend producing diagnostic metrics of coverage over target and amplicon bases, as it can be helpful in troubleshooting issues with coverage. To do this, we recommend CollectTargetedPcrMetrics in the GATK suite, which is built for this purpose

```
java -Xmx4G -jar {path_to_GATK_jar} CollectTargetedPcrMetrics --COVERAGE_CAP 50000
--PER_TARGET_COVERAGE
{per_target_coverage_file.tsv} -TI {target_intervals.bed} -AI
{amplicon_intervals.bed} -I sorted_alignments.bam -O
{metrics_file.tsv} -R {path_to_reference_genome}
```

Where `path_to_GATK_jar` is a path to the JAR file containing GATK, target_intervals.bed is the Twist-provided file of target sites on the human genome, amplicon_intervals.bed is the Twist-provided file of amplicon intervals on the human genome, and path_to_reference_genome is a path to a FASTA containing the human reference genome used for alignment. This step will produce two files—a per-target coverage file with read depth over the SNP sites, and a summary metrics file containing metrics that assess the quality of the targeted sequencing. A detailed description of these metrics can be obtained from the Broad Institute's website (https://gatk.broadinstitute.org/hc/en-us/articles/5257864518811-TargetedPcrMetrics).

CALL GERMLINE VARIANTS (GATK4/HAPLOTYPECALLER)

After aligning to the reference genome, the next step is to call germline variants at the target positions used in the product. We recommend HaplotypeCaller from GATK for this step. An example command is as follows:

```
java -Xmx4G -jar {path_to_GATK_jar} HaplotypeCaller -I sorted_alignments.bam -O sorted_
alignments.vcf -R {path_to_reference_genome} -L {target_intervals.bed}
```

DETERMINE CONCORDANCE BETWEEN SAMPLES

The final step is to determine concordance between the amplicon sample analyzed here, and the matched hybrid capture sample. First, we'll select out only variants in the hybrid capture sample that match the target sites of the SampleID amplicons

```
java -Xmx4G -jar {path_to_GATK_jar} SelectVariants -V {hybrid_capture_sample.vcf} -O
{filtered_hybrid_capture_sample.vcf} -L {target_intervals.bed}
```

Finally, we'll compare calls in the SampleID set to calls from the hybrid capture sample at the same sites. We'll do this with the GATK tool concordance, using the command shown below

```
java -Xmx4G -jar {path_to_GATK_jar} GenotypeConcordance --TRUTH_VCF {filtered_hybrid_
capture_sample.vcf} --CALL_VCF sorted_alignments.vcf -O {concordance_report_sample}
```

This will report a table with an output similar to the table below into a file with the extension "`genotype_concordance_contingency_metrics`"

VARIANT_TYPE	TRUTH_SAMPLE	CALL_SAMPLE	TP_COUNT	TN_COUNT	FP_COUNT	FN_COUNT	EMPTY_COUNT
SNP	Capture_sample	Amplicon_sample	19	14	0	0	0
INDEL	Capture_sample	Amplicon_sample	0	0	0	0	0

If two samples are identical, we should see 100% of the calls fall into the "TP" and "TN" categories (for true-positive and true-negative). Any FP calls would indicate a non-reference call present in the amplicon sample but not the capture sample, and any FN calls would indicate a non-reference call from the capture sample not present in the amplicon sample.