Swift BIOSCIENCES™

Incorporating Molecular ID Technology

Accel-NGS® 2S MID Indexing Kits

Molecular Identifiers (MIDs)

- MIDs are indices used to label unique library molecules
- MIDs can assess duplicate molecules in sequencing data
  - Remove amplification errors and sequencing errors
  - Retain unique molecules identified as duplicates by bioinformatics
- Applications that can benefit from MIDs
  - Low frequency variant detection
    - Hybridization capture with FFPE or cell-free DNA (cfDNA)
  - De-duplication from single read sequencing
    - ChIP-Seq
  - De-duplication from non-random fragmentation
    - cfDNA

Swift BIOSCIENCES™

Applications That Can Benefit from MIDs

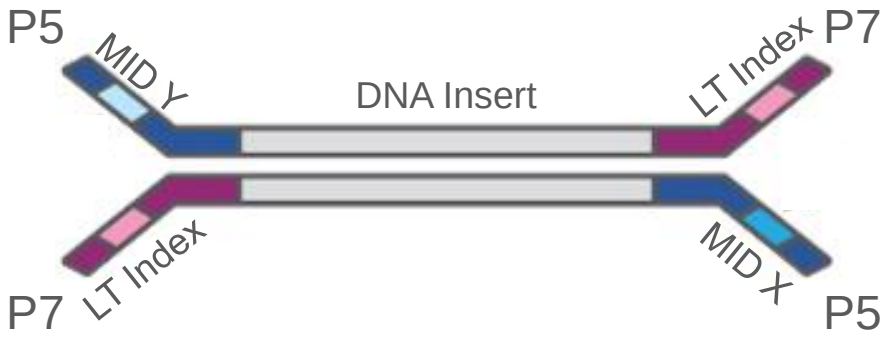
Common sources of duplicates include:

- **Fragmentation duplicates**
  - Inserts with termini that have identical genomic coordinates due to non-random fragmentation
  - Can be retained using MIDs
- **PCR duplicates**
  - Library amplification by PCR generates copies of unique library molecules
  - Can be accurately identified using MIDs
- **Strand partner duplicates**
  - Sense and antisense strands from the same dsDNA insert
  - Can be retained using MIDs
- **Optical/clustering duplicates**
  - Arise on the flow cell by incorrectly identifying a single cluster as two independent clusters
  - Can be identified using cluster coordinates

Swift BIOSCIENCES™

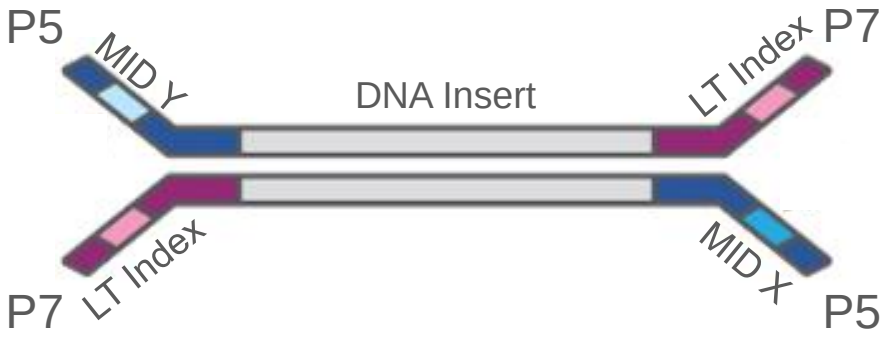
The Structure of Swift MID Libraries

- MIDs are random indices that are used to label unique library molecules
- Swift MID technology is compatible with Accel-NGS® 2S DNA Library Kits
  - P5 adapter sequence: TruSeq® HT with a 9 base random N sequence at the Index 2 position
  - P7 adapter sequence: TruSeq LT with a 6 base standard index at the Index 1 position



Swift BIOSCIENCES™

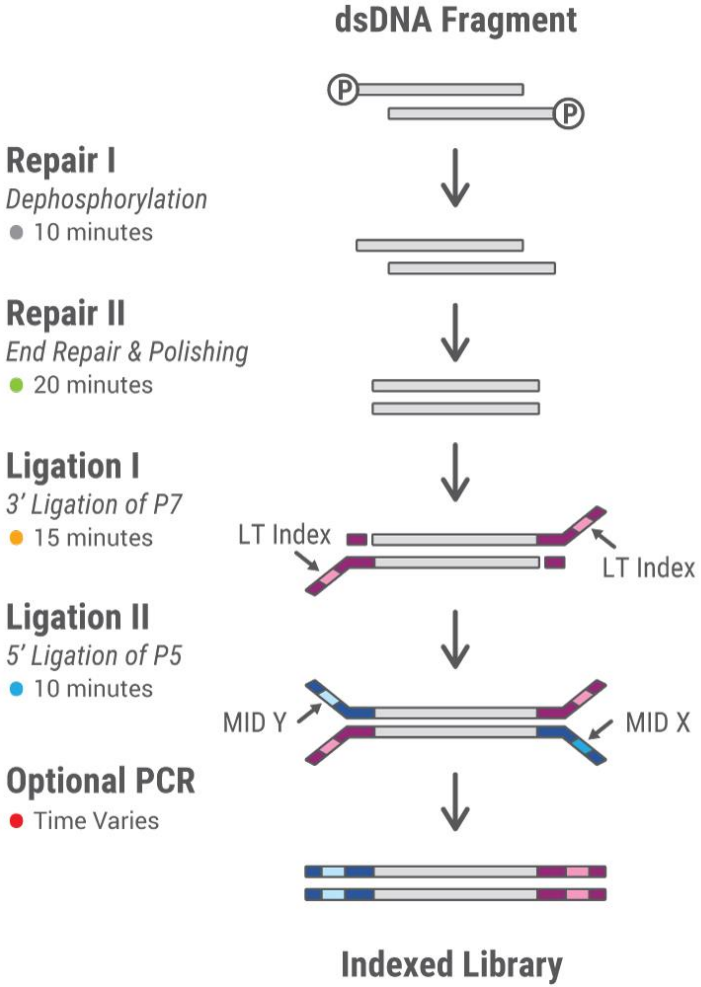
Swift Molecular IDs Are Strand-Specific



Each dsDNA substrate receives two independent P5 MID adapters ‘X and Y’

Due to the polishing reaction, both MID families X and Y will have the same insert.

Using MIDs with Accel-NGS® 2S Kits



Important guidelines for using 2S MIDs:

1. Use the P5 MID adapter (**Reagent B2-MID**) during the Ligation II Step.

2. Perform a **dual Post-Ligation II SPRI cleanup** to ensure optimal removal of unincorporated MID adapter (see protocol for details).

3. If performing hybridization capture with the SureSelect system, **be sure to use the 2S SureSelect^{XT} MID Compatibility Module** (further details on slides 13-15) .

4. Set up your sequencing run to **produce a fastq file from the Index 2 read**, so that you can obtain and analyze the MID sequencing (further details can be found in the “Run Setup and Bioinformatic Analysis” presentation).

Swift BIOSCIENCES™

Accel-NGS® 2S MID Validation

- MID libraries were prepared with Accel-NGS 2S DNA Library Kits.
- A 9 base random N sequence MID has 262,144 theoretical MID sequence combinations.
- Both PCR-free and amplified 2S libraries attain close to theoretical representation of MID combinations with high copy uniformity (percentage of MIDs that are within 20% of the average MID copy number).

Library	Unique MIDs	Mean MID Copy #	MID Copy Uniformity	Max Copy Number	Max Copy MID	Total Index 2 (I5) Reads
PCR-Free	262,104	20.98	98%	464	GGTTACTGG	5498922
9 PCR Cycles	262,036	18.50	99%	447	ACTCTTTCC	4848690

- MiSeq® instruments that are maintained according to Illumina® recommendations typically have sample carryover rates at or below 0.1% (one read in a thousand).
- May need to subtract out artifacts with Illumina TruSeq® P5 adapter sequences from the previous sequence run due to this fluidics contamination.
  - This can be represented by an over-representation of the MID sequence “TCTTTCCCT”, which is the sequence of the Illumina P5 universal adapter at the Index 2 location.
- Can reduce this by using more stringent Illumina sanctioned washes.

Swift BIOSCIENCES™

Applications That Can Benefit from MIDs

Low Frequency Variant Detection (e.g., hybridization capture with FFPE or cell-free DNA)

An MID sequence can be used to distinguish between low frequency mutations and artificial mutations generated either by polymerase errors during PCR amplification or sequencing errors:

- For polymerase errors occurring in the first cycle of PCR, the artificial mutation would be expected to be present in 50% of the PCR duplicate library molecules associated with the same MID.
  - Errors introduced at later PCR cycles would be represented by frequencies less than 50%.
- Sequencing errors would be expected to be present in only one duplicate library molecule for each MID sequence.

- A true low frequency mutation present in the original template could be identified by its presence in greater than 50% of PCR duplicate library molecules with the same MID.
  - Therefore, identification of low frequency mutations requires at least 3 PCR duplicate library molecules for each MID sequence, which corresponds to sequencing your sample to near saturation.

Continued…

Swift BIOSCIENCES™

Applications That Can Benefit from MIDs

Low Frequency Variant Detection (e.g., hybridization capture with FFPE or cell-free DNA)

We recommend working within the following parameters to be cost-effective:

- 10-50 ng of input DNA
  - For 10 ng human samples with a mutation at a frequency of 0.1%, the mutation would be expected to be present in just 3 genomes out of the 3000 genomes in the sample. Using less than 10 ng will impact the ability to detect a 0.1% mutation.
  - Inputs greater than 50 ng are expected to contain a large number of unique library molecules, and require more reads to achieve the depth required for a minimum of 3 duplicates per MID.
- A targeted panel size of 80-800 kb
  - Sequencing targets greater than 800 kb will require more reads to achieve near saturation.

Swift BIOSCIENCES™

Applications That Can Benefit from MIDs

De-duplication from single read sequencing (e.g., ChIP-Seq)

Commonly sequenced with a single read, ChIP-Seq data uses the genomic coordinate of only one end of the insert to remove PCR duplicates. An MID can be used to distinguish fragment duplicates from PCR duplicates with a common genomic coordinate. This prevents fragment duplicates from being removed during standard de-duplication, which preserves library complexity.

De-duplication from non-random fragmentation (e.g., cell-free DNA)

Some cell-free DNA fragments may be generated at non-random positions since nucleosomal positioning is not random, which could produce fragmentation duplicates in your library. By using aligned map position combined with MIDs, PCR duplicates can be distinguished from fragment duplicates. Standard de-duplication tools alone would eliminate fragment duplicates, lowering actual library complexity.

Swift BIOSCIENCES™

Swift Accel-NGS® 2S Hyb + MIDs with SureSelect^{XT}

- The 2S SureSelect^{XT} MID Compatibility Module for MIDs includes non-indexed adapters, and pre-hybridization PCR primers.
  - Ligation I adds a non-indexed, truncated P7 adapter.
  - Ligation II add a full length, MID-containing P5 adapter.
  - Non-indexed libraries cannot be multiplexed in the SureSelect^{XT} hybridization capture reaction.
  - To ensure that the MID sequence on the P5 MID adapter is retained, **these pre-hybridization PCR primers must be used**.
  - Post-hybridization primers (provided by Agilent) can be used after the hybridization capture to add index sequence to the library molecules.

Swift BIOSCIENCES™

Swift Accel-NGS® 2S Hyb + MIDs with SureSelect^{XT2}

- The 2S SureSelect^{XT} MID Compatibility Module for MIDs includes non-indexed adapters, and pre-hybridization PCR primers.
  - Ligation I adds a non-indexed, truncated P7 adapter.
  - Ligation II add a full length, MID-containing P5 adapter.
  - Pre-hybridization custom primers (supplied by user, used in place of pre-hybridization PCR primers in module) will add an 8 bp index sequence to P7 and complete the library molecules. Contact Technical Support for primer recommendations.
  - Indexed libraries can be multiplexed in the SureSelect^{XT2} hybridization capture reaction.
  - Post-hybridization primers (provided by Agilent) can be used after the hybridization capture to amplify the library.

Swift BIOSCIENCES™

Is an MID-Specific, Custom Blocker Needed?

- For Agilent SureSelect^{XT}, we found no benefit to spiking in a custom MID adapter blocker with the standard Agilent blocker.
  - Off target hybridization was the same when using a custom blocker or a standard blocker
  - If desired, a custom TruSeq® HT universal blocker with a 9 base insertion interval can be ordered through IDT custom synthesis.

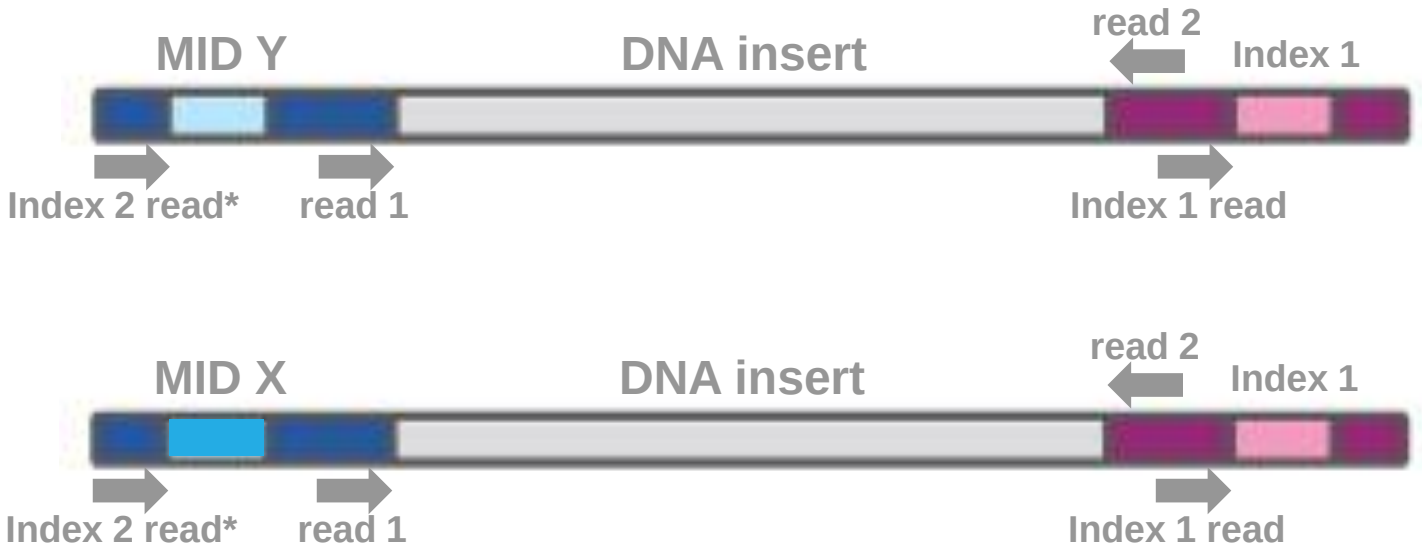
Sample	Input Quantity	Total Read Pairs	Reads Aligned	Median Insert Size	Duplicates	Estimated Library Size	Off Bait	On Bait Vs Selected/ On-Target (bases)	Mean Bait Coverage	Covered 10X	Covered 20X
NA12878	25 ng	34,665,204	97.6%	217 bp	6.0%	200,204,984	**10.15%**	71.76%	40X	95.73%	82.31%
NA12878	25 ng	36,928,298	97.7%	243 bp	4.9%	261,178,502	**10.17%**	68.65%	43X	96.11%	83.77%

Accel-NGS® 2S Hyb DNA Library Kits were used to construct libraries for hybridization capture with the Agilent SureSelect^{XT} V5 Exome Panel. Libraries were sequenced with HiSeq 2500 Rapid Run 100 bp PE reads. On Bait calculations were based on enriched molecules within the target region without any buffer region. On Bait calculations that include buffer would be expected to exhibit a higher percentage.

Swift BIOSCIENCES™

Sequencing MID Libraries

MID X and Y families cluster and sequence independently.



*Index 2 read shown as performed for the MiSeq® and HiSeq® 2500, where this orientation is reversed when sequencing Index 2 on the NextSeq® or HiSeq 3000/4000.

Swift BIOSCIENCES™

Sequencing MID Libraries

For MiSeq®, HiSeq®, and NextSeq® instruments:

- Modify the config file to create a fastq for index reads
- Using the Illumina® Experiment Manager software, specify 2 index reads for the run.
- In the CSV file, specify the MID index and the sample index sequences.
  - Samples will be demultiplexed based only on their sample index.
- Using a custom script, join MIDs to their respective fastq read headers, align these fastq, and analyze the reads with a common genomic coordinate.
- **Further details can be found in the "Run Setup and Bioinformatic Analysis" presentation.**

Swift BIOSCIENCES™

Logic for Bioinformatic Analysis of ChIP-Seq and cfDNA

- After the sequencing run, all reads will be separated based on Index 1 only. Non-identified (or undetermined) reads are either due to poor quality or reads missing index sequences.
- Using a custom script, join the MIDs to the respective fastq read headers.
- Align these fastq using BWA.
- Using another custom script, analyze the reads with a common genomic coordinate.
  - If the reads have unique MID sequences, they represent fragment duplicates and should be retained as unique reads
  - If the reads have identical MID sequences, they represent PCR duplicates. Mark all reads except one as duplicates based on mapping and base quality.

Swift BIOSCIENCES™

Logic for Bioinformatic Analysis of Low Frequency Variants

- After the sequencing run, all reads will be separated based on Index 1 only. Non-identified (or undetermined) reads are either due to poor quality or reads missing index sequences.
- Using a custom script, join the MIDs to the respective fastq read headers.
- Align these fastq using BWA.
- Use Picard Tools or Samtools to collect metrics like % duplication, reads on target, etc.
- Using another custom script, group fragments with at least 3 PCR duplicate reads per MID for variant calling (also considering Flag and Cigar values). Within the group, determine a consensus sequence for each fragment which eliminates sequencing and PCR errors present at less than 50%.
  - The following publication contains some details concerning this kind of analysis, and may be a useful reference:
    - S.R. Kennedy, et al. Nat Protoc. 2014 November; 9(11):2586-2606

Swift BIOSCIENCES™

Ordering Information

Catalog No.	Description	Reactions
27148	2S Set A MID Indexing Kit (12 indices)	48
27248	2S Set B MID Indexing Kit (12 indices)	48
27396	2S Set A+B MID Indexing Kit (24 indices)	96
27424	2S SureSelect^{XT} MID Compatibility Module	24
27496	2S SureSelect^{XT} MID Compatibility Module	96

Click to discover more about the Accel-NGS® 2S MID Indexing Kits.

Swift BIOSCIENCES™

THANK YOU

www.swiftbiosci.com

© 2016, Swift Biosciences, Inc. The Swift logo is a trademark and Accel-NGS is a registered trademark of Swift Biosciences. Illumina, TruSeq, MiSeq, HiSeq, and NextSeq are registered trademarks of Illumina, Inc. SureSelect^{XT} is a product of Agilent Technologies. 16-1168, 10/16

Swift BIOSCIENCES™