Roche

NGS Adapters

What they are, what they do, and how to choose them

Table of Contents

A brief introduction to sequencing adapters and library preparation ... 3
- NGS library preparation: A simple view **(Box 1)** ... 3

A deeper exploration of adapter structure and function ... 4
- Functional motifs of adapters **(Box 2)** ... 4
- Structural types of adapters **(Box 3)** ... 5
- Library prep workflows differ for full-length and truncated adapters **(Box 4)** ... 5
- Summary: Full-length vs truncated adapters **(Box 5)** ... 6

Unique molecular identifiers (UMIs): What they are and how they can increase the accuracy and sensitivity of NGS data ... 7
- Advantages of using UMIs in NGS workflows **(Box 6)** ... 7
- UMIs can increase accuracy of quantitative sequencing data **(Box 7)** ... 8
- UMIs can increase sensitivity and confidence in variant calling **(Box 8)** ... 9
- Summary: Sequencing indexes vs UMIs **(Box 9)** ... 9

Index hopping (index mis-assignment): What it is and how it happens ... 10
- Steps where index mis-assignment can occur **(Box 10)** ... 10

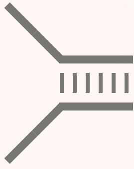
Indexing schemes: What they are and why they matter ... 11
- Dual-indexed (DI) and unique-dual indexed (UDI) indexing schemes ... 11
- DI vs UDI adapters: An overview **(Box 11)** ... 11
- Unique dual-indexed UDI adapters mitigate the impact of index hopping ... 12
- The potential impact of index hopping on NGS data **(Box 12)** ... 12
- How UDIs can mitigate the impact of index hopping **(Box 13)** ... 13

A big-picture summary of library molecule structure (Box 14) ... 14

Factors to consider when choosing an indexing scheme ... 15

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

A brief introduction to sequencing adapters and library preparation



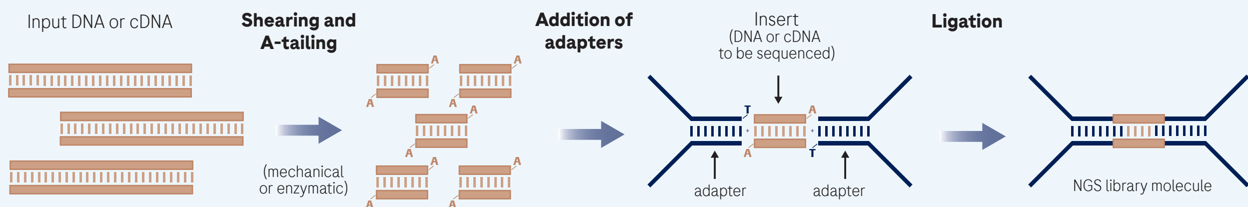
***Sequencing adapters** (NGS adapters) are short DNA sequences that are attached to double-stranded DNA or cDNA fragments to prepare them for next-generation sequencing (NGS).*

NGS adapters come in several varieties. However, common functions of all NGS adapters used to create ligation-based libraries for sequencing-by-synthesis (SBS) technologies, such as Illumina and Singular Genomics sequencing, include:

1. **Converting input DNA (or cDNA) into library molecules** in preparation for sequencing **(known as library construction or library preparation; see Box 1**), and
2. **Enabling sample pooling (multiplexing)** to conserve sequencing resources by including sample-specific index sequences in the adapters.

Box 1 | NGS library preparation: A simple view

using full-length, Y-shaped adapters as an example



Simplified overview of library preparation—the conversion of input DNA to library molecules by the ligation of NGS adapters. During library preparation, the fragmented DNA insert undergoes a 3’ A-tailing reaction prior to ligation to the sequencing adapter, which contains a T overhang. Adapter molecules are then ligated to the inserts, forming library molecules.

Note: While *library preparation* and *library construction* are terms for the entire process of creating the completed, sequencing-ready libraries, the initial process of adding the adapters to the target sequences (DNA inserts) is known as *library conversion*. The *conversion rate* is a measure of the efficiency of this process under a given set of conditions, and is impacted by many factors.

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

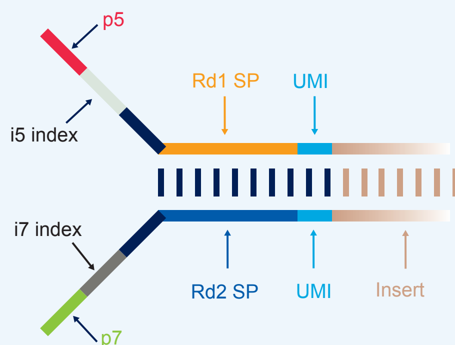
A deeper exploration of adapter structure and function

Box 2 | Functional sequence motifs of adapters

*The diagram below uses a **full-length adapter** (also known as a Y-shaped adapter) to illustrate the main functional components of adapter molecules* **(see Box 3 for a review of adapter types)**.*

Because full-length adapters include all of the motifs necessary for sequencing and demultiplexing, these adapters are well-suited for PCR-free workflows (such as PCR-free WGS), and can also be used in workflows where PCR is required (target-enriched NGS, RNA-seq, etc).

i5 and i7 indexes are included in most sequencing adapters**. These are sample-specific sequence tags assigned to each library; every sequencing read from a given sample will share the same index, or pair of indexes. Indexed adapters allow sample libraries to be pooled (multiplexed) and sequenced together; the resulting sequencing reads can then be sorted and assigned to the correct sample. Thus, the use of indexed adapters can increase the efficiency and reduce the costs of NGS by allowing tens or hundreds of samples to be run in the same sequencing lane or flow cell, as long as each is labeled with a different index or combination of indexes.



p5 and **p7** flow cell oligo binding site sequences are short, invariable sequences that are complementary to the platform-specific flow cell oligos; these motifs enable the library molecules to bind to the flow cell surface on the sequencing instrument.

Rd1 SP and **Rd2 SP** (the Read 1 and Read 2 sequencing primer binding sites) enable the binding of the sequencing primers to the library molecules, which is necessary for the sequencing-by-synthesis (SBS) process to begin.

UMIs (unique molecular identifiers) are optional sequences that uniquely tag each molecule within a library. UMIs (also called molecular barcodes) can help improve data accuracy and sequencing sensitivity **(learn more on Pages 7, 8, and 9 and Boxes 6, 7, and 8)**.

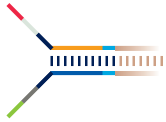
*While these adapters and motifs describe adapters designed for Illumina instruments, adapters for other SBS workflows—such as those used on the Singular Genomics G4 system—function in similar ways.
**Not necessary when no multiplexing (library pooling) will be performed.

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

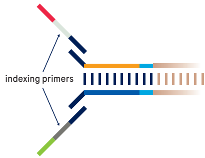
Box 3

Structural types of adapters

There are several different NGS adapter styles, including:



Full-length adapters (Y-shaped adapters), which include all of the sequence motifs required for sequencing and demultiplexing **(see Box 2)**.



Truncated adapters (also known as universal, or stubby, adapters) which typically contain only the sequencing primer binding sites (and may include optional UMIs). Thus, workflows using these adapters require an additional PCR step to incorporate sample indexes and the p5/p7 binding sites, which are included in the **indexing primers (Box 4)**. Although they cannot be used in PCR-free workflows, these adapters provide other advantages **(Box 5)**.

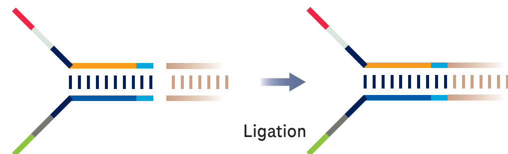
Hairpin adapters, which are specific to certain workflows and require additional enzymatic steps prior to sequencing, as well as the use of PCR to add sample indexes (as shown above*).

Box 4

Library prep workflows differ for full-length and truncated adapters

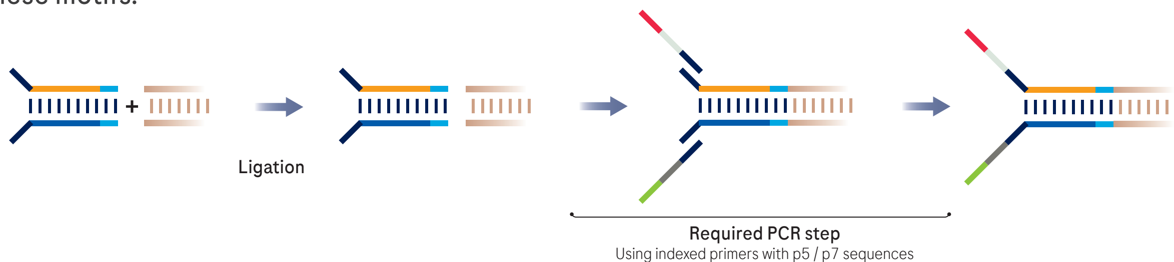
Workflow with full-length adapters

- These Y-shaped adapters are ligated directly onto the double-stranded insert, providing a streamlined workflow.
- Full-length adapters are especially useful for PCR-free workflows.



Workflow with truncated adapters and indexing primers

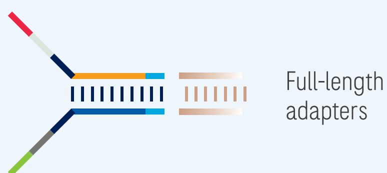
- Truncated adapters contain the sequencing primer binding sites, and may include UMIs, but do not include sample indexes or p5/p7 binding sites.
- The sample indexes and p5/p7 binding sites are added by PCR using indexing primers that contain these motifs.



*In this ebook, we focus on the full-length and truncated adapters described above. However, the functions of each motif are similar across adapter styles.

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

Box 5 | Summary: Full-length vs truncated adapters

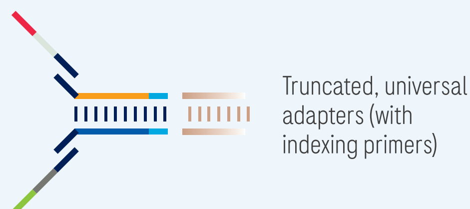


Advantages

- Compatible with PCR-free workflows
- Simpler library preparation workflow
  - Less time
  - Fewer reagents
  - Fewer pipetting steps

Disadvantages

- Fewer options for including UMIs
- May have reduced multiplexing capabilities compared to truncated adapters with indexing primers



Advantages

- Higher ligation efficiency
- Less adapter-dimer formation
- Increased indexing flexibility
- Expanded multiplexing capabilities, as the indexing primers are separate from the adapters
- The inclusion of UMIs is more common with truncated adapters

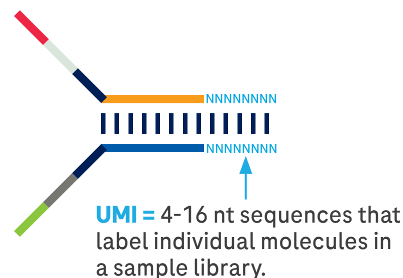
Disadvantages

- Incompatible with PCR-free workflows
- Requirement for additional amplification and cleanup

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

Unique molecular identifiers (UMIs): What they are and how they can increase the accuracy and sensitivity of NGS data

Unique molecular identifiers (UMIs) are short DNA sequences (typically 4 to 16 nucleotides) that are added to each insert molecule during library preparation as part of the NGS adapters, giving each molecule a unique identifying sequence, or barcode **(see Box 2 and below). Other common terms for UMIs include unique identifier (UID), unique molecular index, molecular barcode, random barcode, and molecular counter**.



Box 6 | Advantages of using UMIs in NGS workflows

UMIs can reduce quantitative bias—especially with low-input samples, degraded samples, or in target enrichment workflows.

UMIs enable the identification of PCR duplicates in the sequencing data, making it possible to bioinformatically remove duplicate reads from analysis. This increased accuracy is especially important for quantitative studies where uneven PCR amplification can yield biased results. UMIs also aid the identification and removal of PCR- or sequencing-generated artifacts, preventing their interpretation as true variants **(Boxes 7 and 8)**.

UMI sequences enable grouping of sequencing data into read families, in conjunction with alignment coordinates. These read families represent individual insert molecules, thus allowing for more accurate counting for a better representation of input DNA/cDNA **(Box 7)**.

UMI sequences increase confidence in variant calling by enabling the identification of sequence artifacts that may have been introduced during library preparation, amplification, or sequencing **(Box 8)**.

Some applications where UMIs may be especially valuable:

- Copy number variation (CNV)/copy number alteration (CNA) analysis
- Analysis of relative gene expression (RNA-seq)
- Rare variant analysis
- Ultra-low input applications
- Single-cell RNA-seq (scRNA-seq) applications
- ChIP-seq
- Determining specific variants in mixed samples (ie., heterogeneous tumor tissue or microbiome samples)

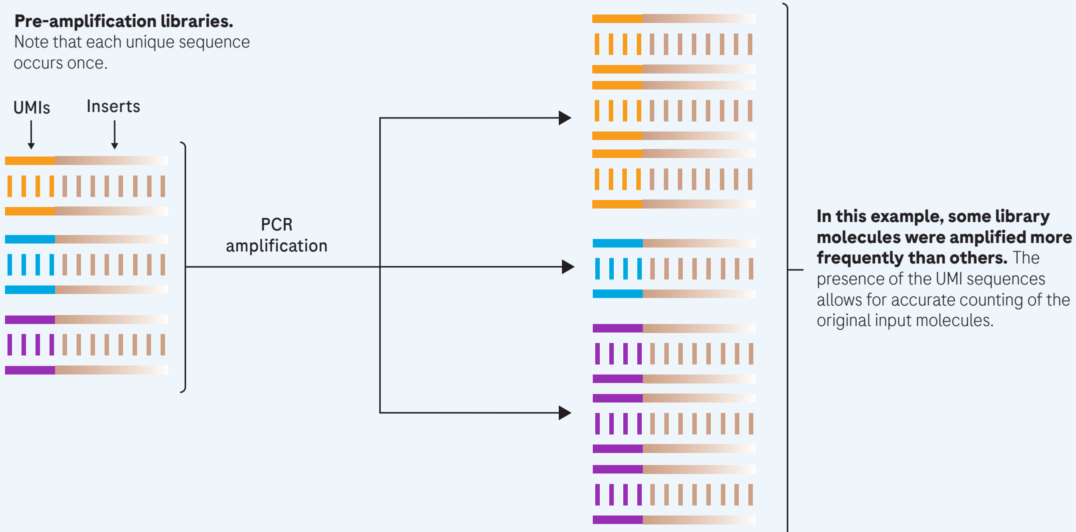
For Research Use Only. Not for use in diagnostic procedures.

Box 7

UMIs can increase the accuracy of quantitative sequencing data

by identifying PCR duplicates... which is especially important when amplification is not uniform.

In the following diagrams **(Boxes 7 and 8)**, the molecules in each box all represent the same genomic region from the same sample library. **The colored segments represent UMIs; for simplicity, index sequences are not shown**. Following sequencing, reads containing the same UMI are considered to be part of the same "read family."



Without UMIs, the integrity of quantitative data may be compromised.

- PCR duplicates may be counted as representing unique input molecules.
- Informative, unique reads that map to the same coordinates may be incorrectly dismissed as being PCR duplicates, causing rare molecules to be missed.

With UMIs, reads can be grouped into read families using bioinformatics.

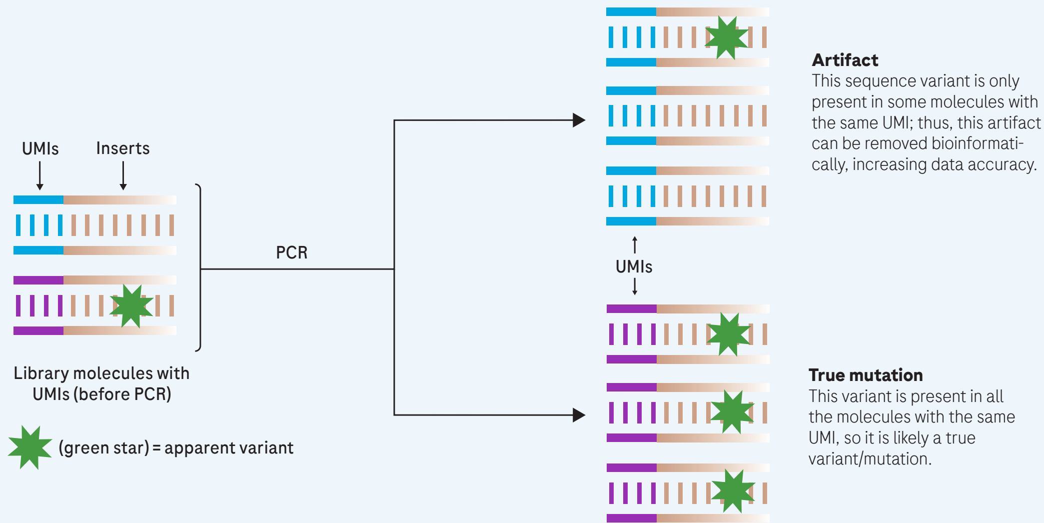
- The correct number of unique input molecules that map to each region can be assessed.
- Relative quantities of input RNA or DNA molecules can be determined more accurately, which is especially relevant for quantitative studies such as differential gene expression or CNV analysis.
- Rare molecules and rare variants are more likely to be identified and/or counted.

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

Box 8

UMIs can increase sensitivity and confidence in variant calling

by enabling the identification of sequencing artifacts in NGS data.



Without UMIs:

- PCR duplicates are not readily identified as duplicates.
- Sequencing artifacts or PCR-induced alterations may be incorrectly counted as true variants.

With UMIs:

- Read families can be identified, so artifacts can be removed from the data... increasing confidence in accurate variant calling.

Box 9

Summary: Sequencing indexes vs UMIs

Sequencing indexes:

- **Provide sample-specific sequence tags** that label each library molecule (and sequencing read) in a given sample library with the same index(es).
- **Enable sample multiplexing and demultiplexing** of sequencing reads, ensuring that the right reads are associated with the right samples.

UMIs:

- **Uniquely tag individual DNA molecules** in an input DNA sample.
- **Increase the accuracy and sensitivity of NGS data** by enabling the identification and removal of PCR duplicates or sequencing artifacts during analysis.

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

Index hopping (index mis-assignment): What it is and how it happens

Index hopping refers to incorrect assignment of sequencing reads to other samples within a multiplexed pool of libraries. This mis-assignment can compromise the integrity of the data and lead to incorrect conclusions about the samples.

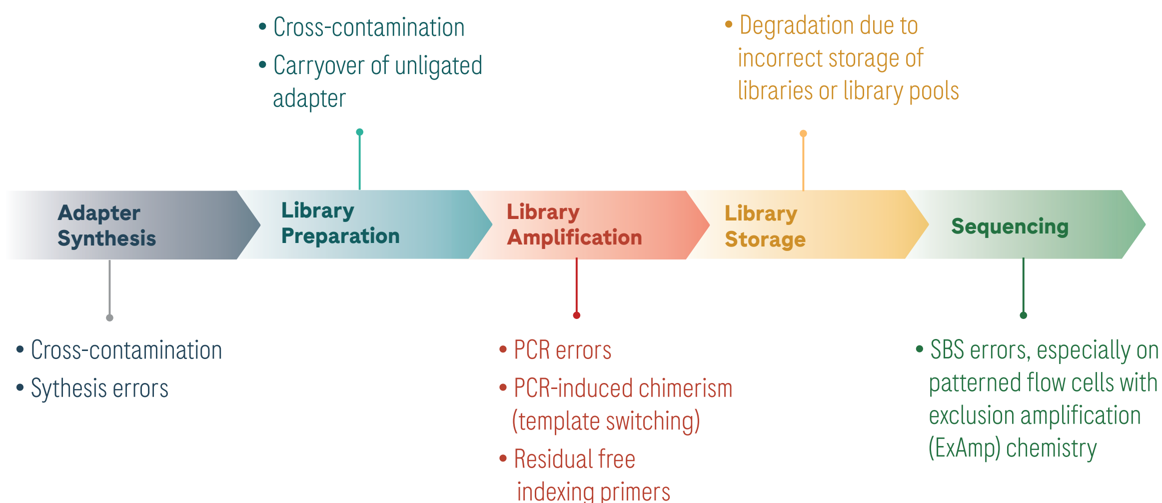
The use of unique dual-indexed adapters can mitigate the impact of index hopping **(see pages 11, 12, 13 and Boxes 11, 12, 13)**.

Other common terms for index mis-assignment:

- index hopping
- index contamination
- index switching
- index swapping
- index cross-talk
- multiplexing noise

Box 10 Steps where index mis-assignment can occur

Index mis-assignment can result from errors at several different steps of the NGS workflow **(see figure below)**, and is especially common when patterned flow cells are used for sequencing.



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

Indexing schemes: What they are and why they matter

Dual-indexed (DI) and unique dual-indexed (UDI) indexing schemes

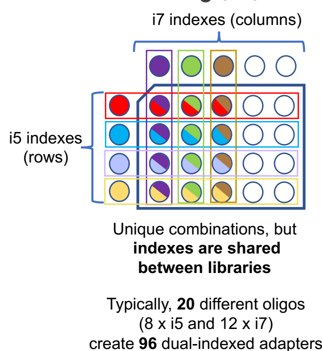
Dual-index adapters are adapters that contain two different indexes (the i5 and i7 indexes are different from each other). These adapters enable the multiplexing of more samples compared to single-index adapters, which are rarely used.* After sequencing, each read will contain either an i5 or i7 index sequence.

DI adapters are often used in **combinatorial indexing schemes;** for example, a dual-indexed adapter kit that includes 8 different i5 adapters and 12 different i7 adapters may be used to create 96 unique index combinations. **In this indexing scheme, while each combination of adapters is unique, some indexes will be shared between samples (Box 11)**. This means that reads from different samples can contain the same index sequence, and care must be taken when deciding on multiplexing schemes.

In contrast, each index sequence in **unique dual-index adapters** is **fully unique**; there is no index redundancy between samples **(Box 11)**. UDI adapters are available as either full-length adapters or as a set of universal adapters and indexing primers. **(See Boxes 3, 4, and 5 for more detail about adapter styles)**.

Box 11 | DI vs UDI adapters: An overview

Dual-indexing (DI) schemes



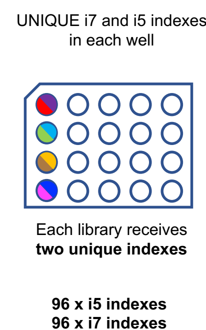
Unique combinations, but **indexes are shared between libraries**

Typically, **20** different oligos (8 x i5 and 12 x i7) create **96** dual-indexed adapters

This image uses a section of a 96-well plate to show how combinatorial indexing schemes work. Here, all wells in a given row contain a specific i5 indexed oligonucleotide (oligo), and the wells in each column contain a specific i7 indexed oligo. Thus, each well yields adapters with a unique combination of indexes, although individual indexes are shared.

Unique dual-indexing (UDI) schemes

UNIQUE i7 and i5 indexes in each well



Each library receives **two unique indexes**

96 x i5 indexes
96 x i7 indexes

This image shows a section of a 96-well plate where every well contains two unique indexes; there are no shared indexes between samples.

DI adapters
(combinatorial)
Samples may share indices, as shown here for the magenta-colored index present in both libraries.

- Can lead to index-hopping
- Reads can be attributed to the wrong samples

UDI adapters
Samples do NOT share indices, as shown by 4 different, non-redundant colors here.

- Allows for filtering (removal) of incorrectly assigned reads caused by index hopping

Here, each box depicts two libraries that were sequenced together: libraries A & B in one run, libraries C & D in another run. The use of UDI adapters ensures that incorrectly indexed samples will be removed from the data. See Boxes 12 and 13 for an illustration.

*Single-indexed adapters may be used when sequencing is being performed on older sequencers not capable of dual-indexing, or where users do not want to perform the additional on-instrument cycles required for a second indexing read.

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

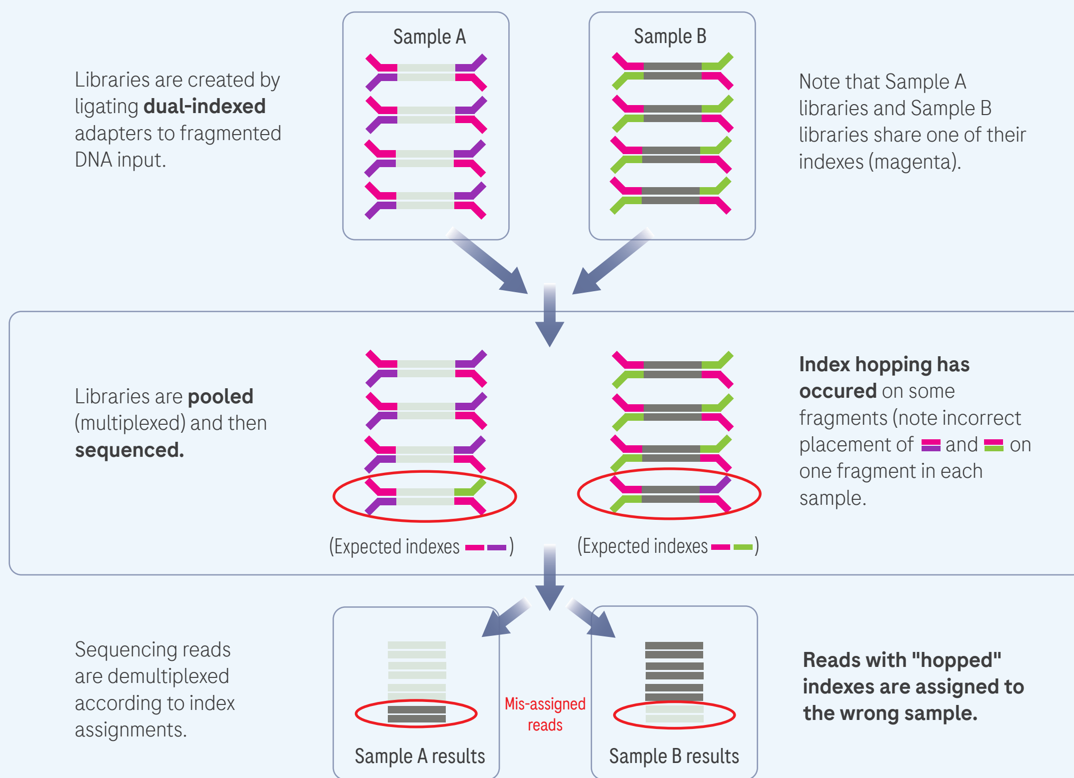
Unique dual-indexed UDI adapters mitigate the impact of index hopping

Unique dual-indexed adapters mitigate the impact of index mis-assignment by enabling the removal of mis-assigned reads, increasing the accuracy of the final results. In Boxes 12 and 13, simplified diagrams of library molecules with full-length adapters are used; although only two NGS libraries are shown in each figure, the same principles apply when many libraries are multiplexed.

Box 12

The potential impact of index hopping on NGS data

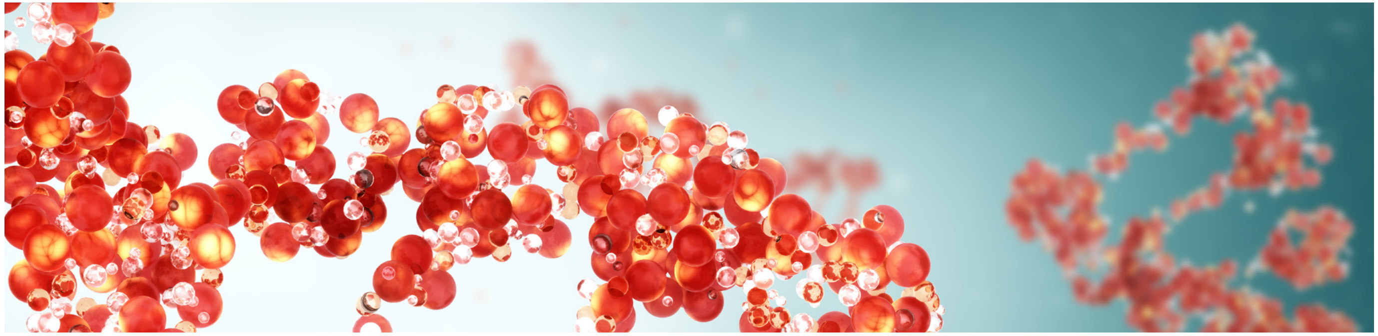
when combinatorial dual-indexing schemes are used



Result: The final data contains SEQUENCING READS FROM OTHER SAMPLES

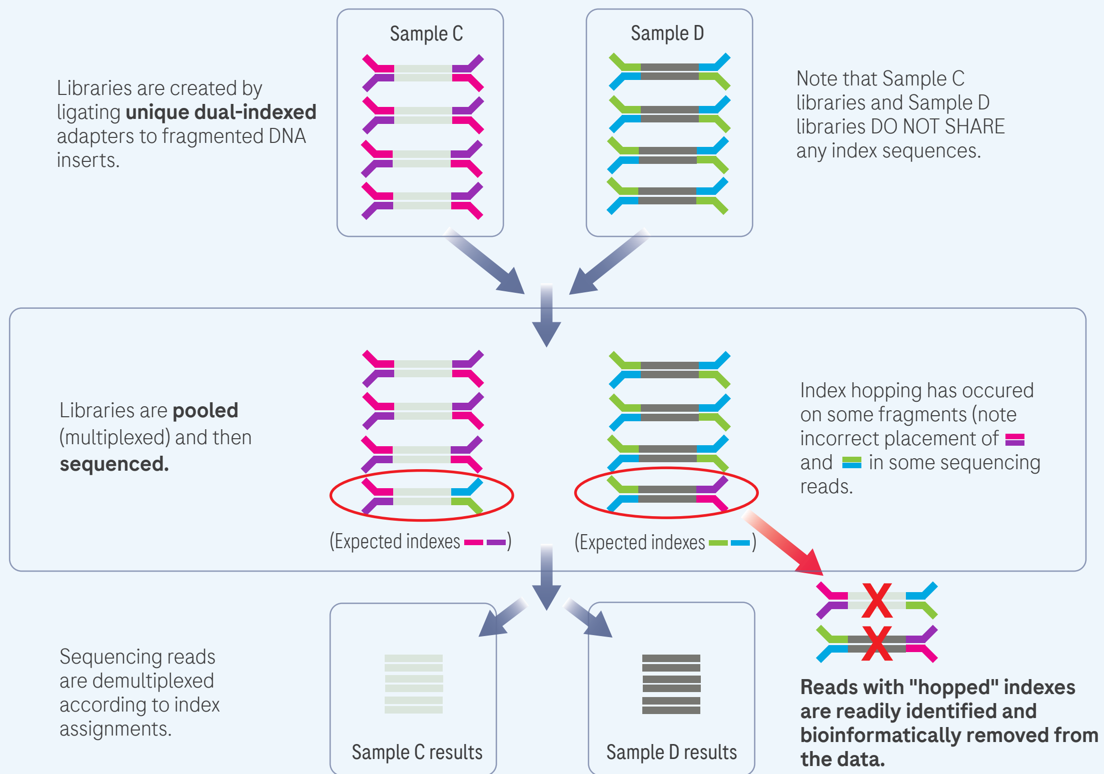
When combinatorial DI adapters are used, some samples may share an i5 or i7 index. If indexes are switched during library preparation, amplification, or sequencing, some sequencing reads can be attributed to the wrong sample during demultiplexing. In this case, some reads may be assigned to the wrong sample.

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



Box 13 | How UDIs can mitigate the impact of index hopping

by enabling the removal of incorrectly indexed NGS reads



Result: The final data contains ONLY CORRECTLY INDEXED READS

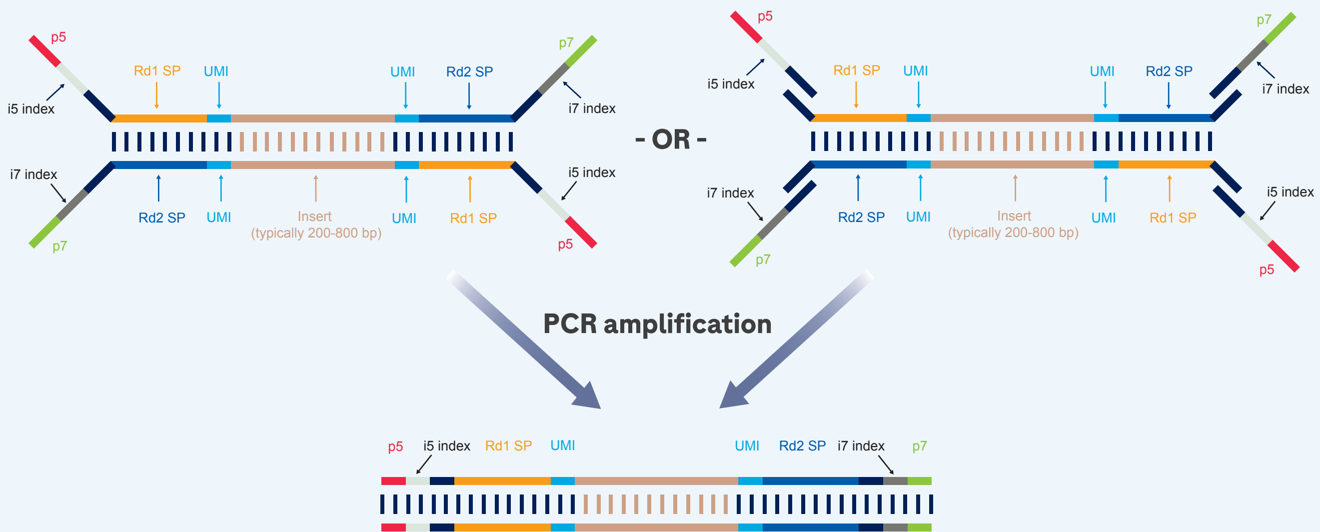
When UDI adapters are used, there is no index redundancy between samples. This enables the bioinformatic identification and removal of reads containing unexpected indexes, leading to increased data integrity.

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

Box 14 | A big-picture summary of library molecule structure

With full-length adapters
ligated directly to inserts
(PCR amplification is optional)

With universal adapters
ligated directly to inserts,
and indexes and p5/p7 sequences
added **with indexing primers** via
subsequent PCR amplification



Following amplification, double-stranded library molecules have a p5 sequence and i5 index on one end, and p7 sequence and i7 index on the other end.

During sequencing, each strand is sequenced separately, starting with a sequencing primer that binds to the Rd1 or Rd2 sequencing binding site.

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

Factors to consider when choosing an indexing scheme

Workflow factors:	**Combinatorial dual-indexed (DI) full-length adapters** are recommended or sufficient	**Unique Dual Index (UDI) adapters** are recommended	**UMIs** might be recommended
You are using a PCR-free workflow	**YES**	**YES** (full-length only)	n/a
Samples will be sequenced on an Illumina instrument with a patterned flow cell	No	**YES**	**YES**
If index hopping due to sequencing workflow is a concern	No	**YES**	**YES**
Identification of PCR duplicates is essential for accuracy	n/a	**YES**	**YES**
Identification of artifacts due to PCR or sequencing errors is critical to data integrity	No	**YES**	**YES**
Identification and/or quantification of rare molecules (cfDNA, ctDNA, rare variants, scRNA-seq, low expressors) is the goal	No	**YES**	**YES**
Accurate quantification is important (copy number variation, ctDNA, relative gene expression, ChIP-seq, relative percentage of molecules in mixed samples	No	**YES**	**YES**
You will be multiplexing large numbers of samples	No	**YES**	**YES**
Recommended Roche products	n/a	**KAPA Universal Adapter** and **KAPA Unique Dual-Indexed (UDI) Primer Mixes** (384 available combinations) OR **Full-length KAPA Unique Dual-Indexed (UDI) Adapters** (up to 96)	**KAPA Universal UMI Adapter** and **KAPA Unique Dual-Indexed (UDI) Primer Mixes** (384 available combinations)

FOR MORE INFORMATION ABOUT NGS ADAPTERS,
contact your Roche sales representative or visit: **go.roche.com/KapaAdapters**

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

Published by:
Roche Sequencing and Life Science
915 Hague Road
Indianapolis, IN 46256

sequencing.roche.com

For Research Use Only. Not for use in diagnostic procedures.
KAPA is a trademark of Roche.
All other product names and trademarks are the property of their respective owners.
© 2023 Roche Sequencing and Life Science. All rights reserved.

MC-US-12986 A544 05/23