Targeted sequencing guide

IDT
INTEGRATED DNA TECHNOLOGIES

YOUR ADVOCATE
FOR THE GENOMICS AGE

www.idtdna.com

Table of contents

1. What is next generation sequencing (NGS)?

Next generation sequencing (NGS), or high-throughput sequencing, enables sequence profiling of everything from genomes and transcriptomes to DNA-protein interactions. These technologies are an integral part of genetic research and discovery. The ability to generate large amounts of sequence data in a relatively short amount of time enables a wide range of genetic analysis applications and accelerates advances in research, clinical, and applied markets.

2. What is targeted NGS?

Targeted NGS allows users to sequence specific areas of the genome for in-depth analyses in a more rapid, cost-effective way than whole genome sequencing (WGS). Targeted sequencing detects known and novel variants within your region of interest. This method generally requires less sample input, produces a smaller amount of data than WGS, and makes analyses more manageable.

There are several methods of targeted sequencing, each appropriate for specific applications. The most popular methods are hybridization capture, amplicon sequencing, and molecular inversion probes (MIPs) (Table 1) [1].

Table 1. Brief comparison of hybridization capture and amplicon sequencing.

Feature	Hybridization capture	Amplicon sequencing
Number of targets per panel	Virtually unlimited	Less than 10,000 amplicons
Applications	Exome sequencing Genotyping Oncology Rare variant detection	Genotyping by sequencing CRISPR genome editing confirmation Detection of disease-associated variants
Sensitivity	Down to 1% without unique molecular identifiers (UMIs)	Down to 5%

2.1 Introduction to hybridization capture

Before to hybridization capture enrichment, samples are converted into sequencing libraries. The DNA is randomly sheared into smaller fragments, sequencing adapters are added, and depending on the library design, PCR-amplified. Regions of interest in this library are then captured using long oligonucleotide baits (Figure 1). Because the DNA was randomly sheared during library preparation, captured fragments are overlapping and unique. Baits can be tiled, overlapped, and positioned to overcome challenges of repetitive sequences, etc. With advanced design, capture can be made very uniform.

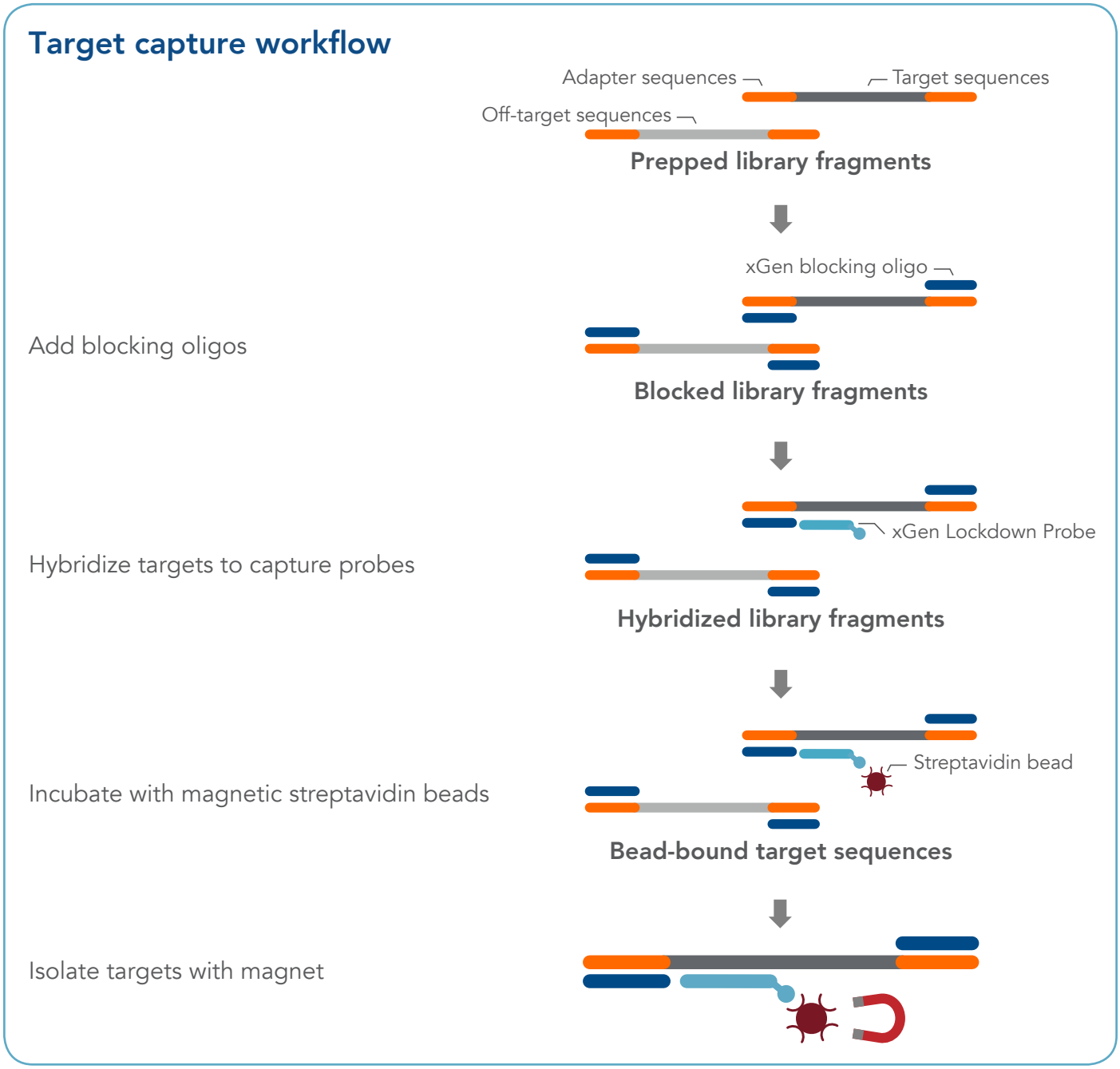

Figure 1. xGen hybridization capture workflow. xGen Blocking Oligos improve enrichment performance by binding to platform-specific adapters to prevent cross-reactivity between library fragments. xGen Lockdown Probes (baits) bind to target regions of interest during in-solution hybrid capture. Targeted regions are then pulled out of solution using streptavidin beads.

2.1.1 Use case: Custom hybridization capture panels are used to detect and sequence Hepatitis C virus genotypes

Hepatitis C virus (HCV) is one of many incurable diseases for which personalized treatment based on genetics would greatly improve patient outcomes. Efficient methods for whole genome viral sequencing offer insights into effective treatment strategies. Using **xGen Lockdown Probes**, Bonsall et al. developed a novel approach for high throughput, whole genome HCV sequencing [2]. Their method, ve-SEQ, improves on conventional viral sequencing methods such as PCR amplification and RNA-Seq based metagenomics.

They constructed indexed sequencing libraries from plasma RNA of HCV-infected patients and performed both conventional RNA-Seq and ve-Seq using hybridization capture. In their ve-SEQ approach, they enriched the previously-sequenced pool of libraries with an optimized capture panel made from xGen Lockdown Probes. The panel contained 4 sets of 155–157 biotinylated probes, each 120 nt, and provided capture of 4 HCV genotypes.

The researchers found that the RNA-Seq-based approach generated 1% total HCV sequence yield from samples with high viral load (VL). They note how this approach, while unbiased, is costly and insensitive to samples with low to mid-range VL. With their ve-SEQ method, however, the group was able to generate linear detection of VLs down to 1000 IU/mL—more than 10X lower than by regular RNA-Seq. In addition, ve-SEQ (with xGen probes) led to a 224X increase in total HCV reads, including 1000X enrichment for samples with low to mid-range VL, and nearly 100% HCV sequence content for high VL samples. With ve-SEQ, near-maximal enrichment is achievable if the sequence of a sample segment and its closest matching probe is ≥80% identical. The researchers implemented an algorithm for optimizing probe design, resulting in the creation of a comprehensive HCV panel representing 6 of the 7 currently recognized HCV genotypes.

The ve-SEQ based, optimized panel ultimately offered robust, unbiased detection of resistance-associated variants (RAVs) in a clinical setting. The researchers successfully monitored viral re-emergence in patients undergoing antiviral drug treatments, and identified known drug resistant alleles within these new populations. Overall, the group's findings support ve-SEQ as a cost-effective improvement over other high throughput approaches to whole genome viral sequencing and suggest similar methods could be applied to other pathogens.

2.2 Introduction to amplicon sequencing

Amplicon sequencing is a highly targeted approach that enables you to analyze genetic variation in specific genomic regions. This method uses PCR to create sequences of DNA called amplicons. The amplicons are indexed and sequenced. A specialized version of amplicon sequencing is based on the use of RNase H (Figure 2). Amplicon sequencing is typically used for variant detection.

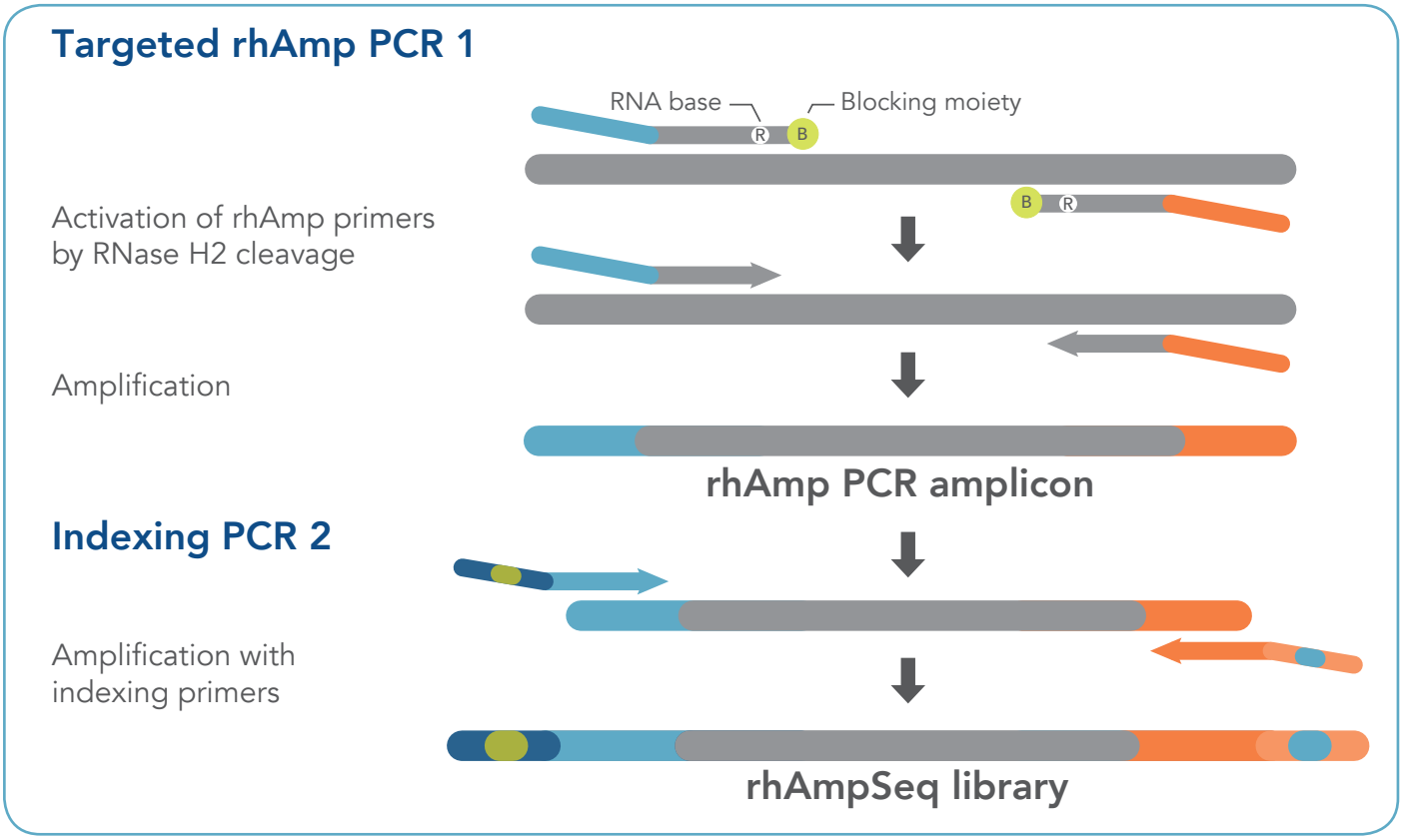

Figure 2. Amplification steps in the rhAmpSeq amplicon sequencing workflow. RNase H2 activates rhAmp primers by target-specific cleavage of the RNA base within the DNA:RNA duplex, removing a 3′ blocker. RNase H2 activity is highly specific, thus reducing the amount of amplification from non-specific hybridization and primer dimers. Only activated rhAmp primers can be extended to generate target amplicons.

2.2.1 Use case: Amplicon sequencing validates genomic markers to map grape diversity

Grape geneticists and breeders use amplicon sequencing to differentiate between grape varieties. Grapes have around 2000 "core genome markers" that can be used to validate the species. These markers are being used to map traits among 6 different populations of grapes in the U.S. The pilot study, funded by the USDA, is called VitisGen2.

Amplicon sequencing allows up to 4000 individual grape DNA samples to be mixed, amplified, and sequenced all at once. The ability to map individual genomes simultaneously is an enormous cost savings.

Professor Bruce Reisch at Cornell University's College of Agriculture and Life Sciences leads a multi-institution research collaboration that adapted rhAmpSeq amplicon sequencing technology to their 250 bp haplotype strategy targeting the core genome. He commented, "As early testers of the rhAmpSeq system from IDT, we were really pleased to see how we were able to rapidly accelerate the VitisGen2 program by using a nearly 2000 marker rhAmpSeq panel to analyze 19 Vitis linkage groups. The core marker set was useful across 6 unrelated populations representing the diversity of the genus, and the workflow was very easy to use and allowed high-throughput processing."

VitisGen2 pioneered the use of this technology in agriculture and helped refine amplicon sequencing with IDT. Read more about how IDT lowers genomic barriers with powerful rhAmpSeq targeted sequencing system.

2.3 Introduction to molecular inversion probes (MIPs)

Molecular inversion probes are another common target enrichment method. Target-specific sequences are ligated to both ends of a universal sequence to make the MIP. Sample DNA can either be fragmented by restriction enzymes, or mechanically sheared. The use of restriction enzymes allows probes to ligate to the restriction sequence. The MIP hybridizes to the region of interest before a gap-filling reaction and a second ligation closes the circles (Figure 3). Target sequences are amplified before sequencing. Molecular inversion probes are primarily used for large-scale genotyping, so the rest of this guide will focus on features of hybridization capture and amplicon sequencing.

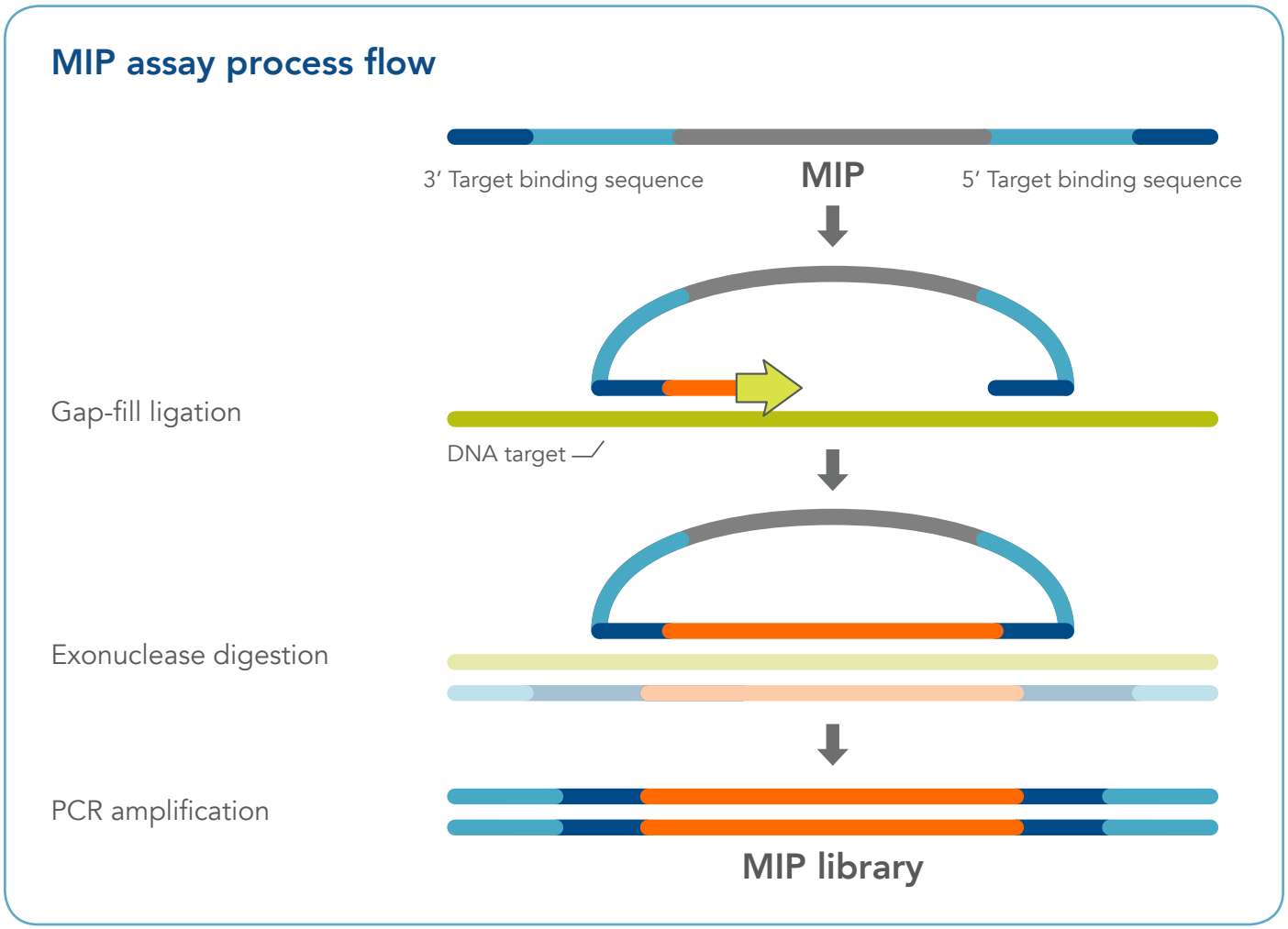

Figure 3. Molecular inversion probe (MIP) workflow. MIPs are comprised of a backbone of a single strand of bases (grey) bordered by universal primer sequences (light blue) and target-binding sequences (dark blue). The target-binding sequences are complementary to your region of interest (green bar). The probe binds to the target regions, the sequence in between is gap-filled using a polymerase (orange), and the circularized probe is ligated closed. The circularized probes are cleaned from the non-target DNA using exonucleases. The single-stranded, circularizied probes containing the target region are amplified using PCR to create the sequencing libraries.

3. Considerations when choosing a targeted sequencing method

3.1 Total time and automation

Amplicon sequencing, which can be completed in less than a day, takes less time than hybridization capture. Some amplicon sequencing protocols have a high-throughput option, which takes the least amount of time. Since hybridization capture requires that the libraries must be prepared before the target enrichment step, additional time is needed. When comparing hands-on time alone, both hybridization capture and amplicon sequencing can take the same amount of time, depending on the number of samples (Table 2).

Table 2. Minimum time requirements for targeted sequencing.

	Typical protocols	IDT protocols
Hybridization capture	4 hours to 4.5 days	4 hours
Amplicon sequencing	5 hours to 1.5 days	1 hour (high-throughput protocol)

While PCR-based approaches (amplicon sequencing) are typically simpler, each step of the protocols for either target enrichment method can be automated on liquid handling robots, with several scripts already existing for hybridization capture. Amplicon sequencing is equally easy to automate.

3.2 Throughput

During amplicon sequencing, library preparation and target enrichment are performed simultaneously. Whereas in hybridization capture, you must first create libraries for each sample, before capture can be performed in parallel by multiplexing. Sample indexes (or indices) enable multiple samples to be sequenced together (i.e., multiplexed) on the same instrument flow cell or chip. Each sample index, typically 6–10 bases, is specific to a given sample library and is used for de-multiplexing during data analysis to assign individual sequence reads to the correct sample. Libraries may contain single or dual sample indexes depending on the number of libraries combined and the level of accuracy desired, but dual indexes are becoming increasingly common [3].

The point at which samples are multiplexed during the workflow is a major differentiator between hybridization capture and amplicon sequencing. Hybridization capture allows multiplexing right after sample libraries are created, so samples can be multiplexed before target capture (enrichment). This feature allows hybridization capture to include more targets per panel. For amplicon sequencing, the targets in each sample library must be enriched separately and the samples can be multiplexed immediately before sequencing.

3.3 Target rate and uniformity

Hybridization capture is liable to produce more off-target reads than amplicon sequencing due to the cross-hybridization of the adapter sequences. The quality of the blockers used to reduce non-specific adapter interaction during probe hybridization can significantly impact the on-target capture performance (See Minimizing duplicates and obtaining uniform coverage in multiplexed target enrichment sequencing). The amount and quality of DNA will impact the on-target rate. Often, increasing coverage depth will increase sensitivity for accurate variant identification. IDT's xGen Lockdown Panels achieve a high on-target performance with uniform coverage that can be achieved with a small amount of DNA.

Amplicon sequencing has naturally higher on-target rates due to the specificity of primer design. The improved on-target rate comes at the cost of decreased uniformity, since an inherent PCR bias occurs when amplifying many targets in a single reaction [4]. Some amplicon sequencing protocols result in "dropouts," missed targets due to a SNP within a primer site preventing the target from being amplified. The quality of rhAmp Primers prevents this dropout, resulting in more uniform coverage.

3.4 Applications

The final factor that should determine the sequencing method used for your experiment is the application. Both methods are able to handle challenging samples, like formalin-fixed paraffin-embedded (FFPE) tissue. Depth of coverage required for your application is an important factor, as sequencing to detect germline mutations requires lower coverage than for somatic mutations, such as those in oncology diagnostic panels.

Hybridization capture works well for genotyping and rare variant detection. Its capacity for mutation discovery makes it particularly suited to oncology research. The sequence complexity and scalability make it a better choice for exome sequencing. If additional sensitivity is needed, unique molecular identifiers (UMIs) can increase the sensitivity of hybridization capture. UMIs are similar to indexes in that they are used for identification. They differ by identifying specific molecules within a sample. This can be used to remove duplicates generated during PCR that may create a bias, which can be particularly helpful when detecting low frequency variants.

Amplicon sequencing can also be used for genotyping by sequencing, for detection of germline SNPs and indels, and known fusions. This method is suitable for applications such as detection of disease-associated variants, as well as validating CRISPR on- and off-target edits after genome editing (See Evaluate CRISPR-Cas9 edits quickly and accurately with rhAmpSeq targeted sequencing). Amplicon sequencing is appropriate for experiments that require less sensitivity, since a bias created during PCR cannot be removed.

4. Conclusion

When you are considering sequencing as part of your research investigations, you must take into account a variety of factors. Targeted sequencing can be a cost-effective choice for many areas of research. Both hybridization capture and amplicon sequencing offer unique options that may be applicable for your research needs. All the factors, such as target rate and uniformity needed for your application, any throughput and time considerations, as well as your automation capabilities, must be taken into consideration when choosing an experimental method that will give you the best results for your application. IDT's xGen Lockdown Panels are a great option for hybridization sequencing and the rhAmpSeq system is a quality selection for amplicon sequencing. Contact an IDT Scientific Application Specialist to help you choose a solution based on your experimental needs.

5. References

1. Kozarewa I, Armisen J, et al. (2015) Overview of target enrichment strategies. Curr Protoc Mol Biol. 112:7.21.21—7.21.23.
2. Bonsall D, Ansari MA, et al. (2015) ve-SEQ: Robust, unbiased enrichment for streamlined detection and whole-genome sequencing of HCV and other highly diverse pathogens. F1000Research. 4:1062.
3. Xie J, Lu X, et al. (2016) Capture-based next-generation sequencing reveals multiple actionable mutations in cancer patients failed in traditional testing. Mol Genet Genomic Med. 4(3):262—272.
4. Kircher M, Sawyer S, et al. (2012) Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res. 40(1):e3.
5. Kebschull JM and Zador AM. (2015) Sources of PCR-induced distortions in high-throughput sequencing data sets. Nucleic Acids Res. 43(21):e143.

Integrated DNA Technologies, Inc. (IDT) is your Advocate for the Genomics Age. For more than 30 years, IDT's innovative tools and solutions for genomics applications have been driving advances that inspire scientists to dream big and achieve their next breakthroughs. IDT develops, manufactures, and markets nucleic acid products that support the life sciences industry in the areas of academic and commercial research, agriculture, medical diagnostics, and pharmaceutical development. We have a global reach with personalized customer service. See what more we can do for **you** at www.idtdna.com.

Technical support:

applicationsupport@idtdna.com

IDT
INTEGRATED DNA TECHNOLOGIES