Targeted sequencing

An introduction to hybridization capture and amplicon sequencing

> SEE WHAT MORE WE CAN DO FOR **YOU** AT **WWW.IDTDNA.COM**.

IDT
INTEGRATED DNA TECHNOLOGIES

custom oligos • next generation sequencing • CRISPR genome editing • qPCR & PCR • synthetic biology • functional genomics

For Research Use Only. Not for use in diagnostic procedures. | Version 1

Table of contents

WHAT IS TARGETED NGS?

Targeted next generation sequencing, or NGS, is a method for sequencing specific areas of a whole genome or specific sequences for in-depth analysis and a more cost-effective method than whole genome sequencing (WGS). In addition, targeted sequencing provides significantly deeper coverage depth to identify low frequency variants, or other rare molecules within a sample. Beneficially, targeted sequencing can identify known and novel variants within a region of interest. Another advantage is that targeting methods generally produce a smaller amount of data than WGS, which makes both sequencing and analysis more manageable.

The two most popular methods for targeted NGS are hybridization capture and amplicon sequencing. Both are appropriate for a wide range of applications. There are also an equally wide range of products available as 'off-the-shelf' to accommodate these research needs, and customization is available for both hybridization capture and amplicon sequencing.

When deciding between the two methods, consider that hybridization capture is best suited for larger, megabase range panels up to and including the whole exome—often for target discovery—whereas amplicon panels are better suited for focused, kilobase range panels that handle smaller sets of target regions of interest.

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

Feature	Hybridization capture	Amplicon sequencing
Number of targets per panel	Virtually unlimited	Flexible, often less than 10,000 amplicons
	Optimization and panel additions are possible	
Applications	• Exome sequencing • Genotyping • Mutation discovery • Oncology research • Rare variant identification	• Oncology research • Genotyping • CRISPR genome editing confirmation • Identification of specific variants • Virology • Metagenomics • Rare variant identification

Introduction to hybridization capture

To conduct a hybrid capture enrichment experiment, you first need to convert your sample into a sequenceable library. If your input sample is not cell-free DNA (cfDNA), it would need to be randomly sheared into smaller fragments, ligated with adapters compatible with your sequencing platform, and PCR-amplified if applicable. Shearing can occur via mechanical, enzymatic, or tagmentation methods. Once your library is made, you can then capture your regions of interest using 5' biotinylated oligonucleotide probes (Figure 1). Randomly shearing your input DNA sample ensures the captured fragments are unique and non-overlapping. For targets of interest that may be more difficult to capture (e.g., repetitive regions), probes can be tiled, overlapped, or placed specifically.

Note: IDT provides an online design tool to help design your targeted NGS research. Our application team is also available to help with your experimental design options by filling out a quick and easy service form here.

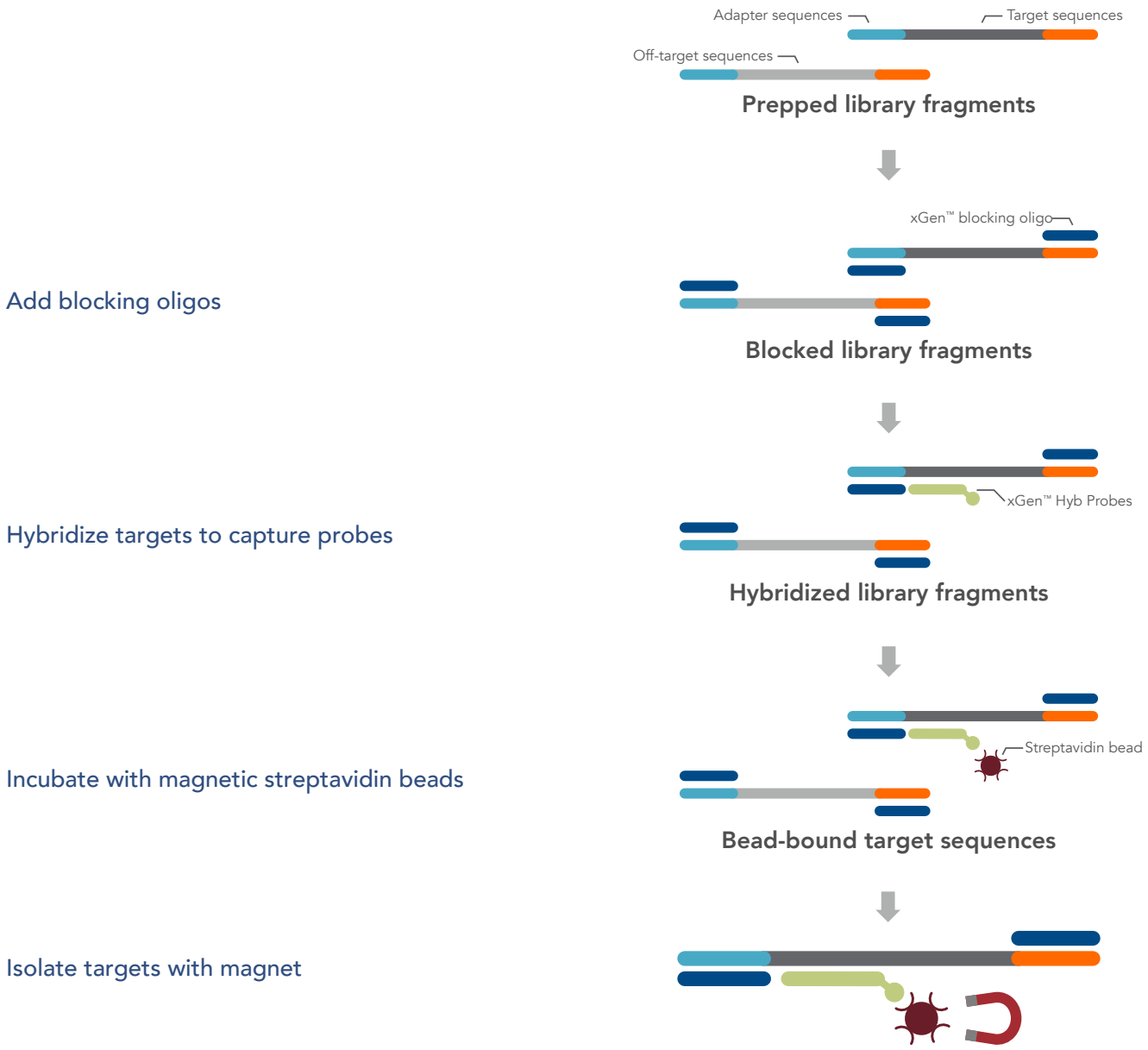

Figure 1. The xGen™ Hybridization Capture overview.

In Figure 1, blocking oligos and Cot DNA are added to your prepped library. Blockers are important for drastically reducing non-specific adapter interactions and Cot DNA reduces non-specific binding of non-coding repetitive sequences during hybridization. If blockers and/or Cot DNA are absent in your reaction, your sample will have low on-target. During hybridization, 5′ biotin oligonucleotides are hybridized to the targets. Typically, this is a four-hour incubation, but can be extended to overnight to improve small panel performance or even acclimate lab schedules. After hybridization completes, magnetic streptavidin beads are added to the reaction to capture probes. Probes are pulled down by magnetic separation racks and a series of washes that take place to wash away any off-target molecules.

Tip: To obtain enough yield for sequencing, you may need to amplify your capture.

Introduction to amplicon sequencing

Amplicon sequencing is a highly targeted approach that enables a deeper analysis of genetic variation in specific genomic regions. With amplicon sequencing, up to hundreds of genes can be sequenced simultaneously.

This method involves a multiplexed PCR approach to create amplicon DNA molecules, which are generated from target specific primer pairs. The amplicons are then converted into libraries with NGS adapters, including sample specific indexes for multiplexed sequencing. Amplicon sequencing has many research uses, including but not limited to:

- Genotyping
- Oncology research (somatic variant identification)
- Metagenomics
- Virology
- Rare variant identification
- CRISPR gene editing confirmation

Amplicon technology differs from hybridization capture by having a shorter workflow (Figure 2), and performing better with smaller panels.

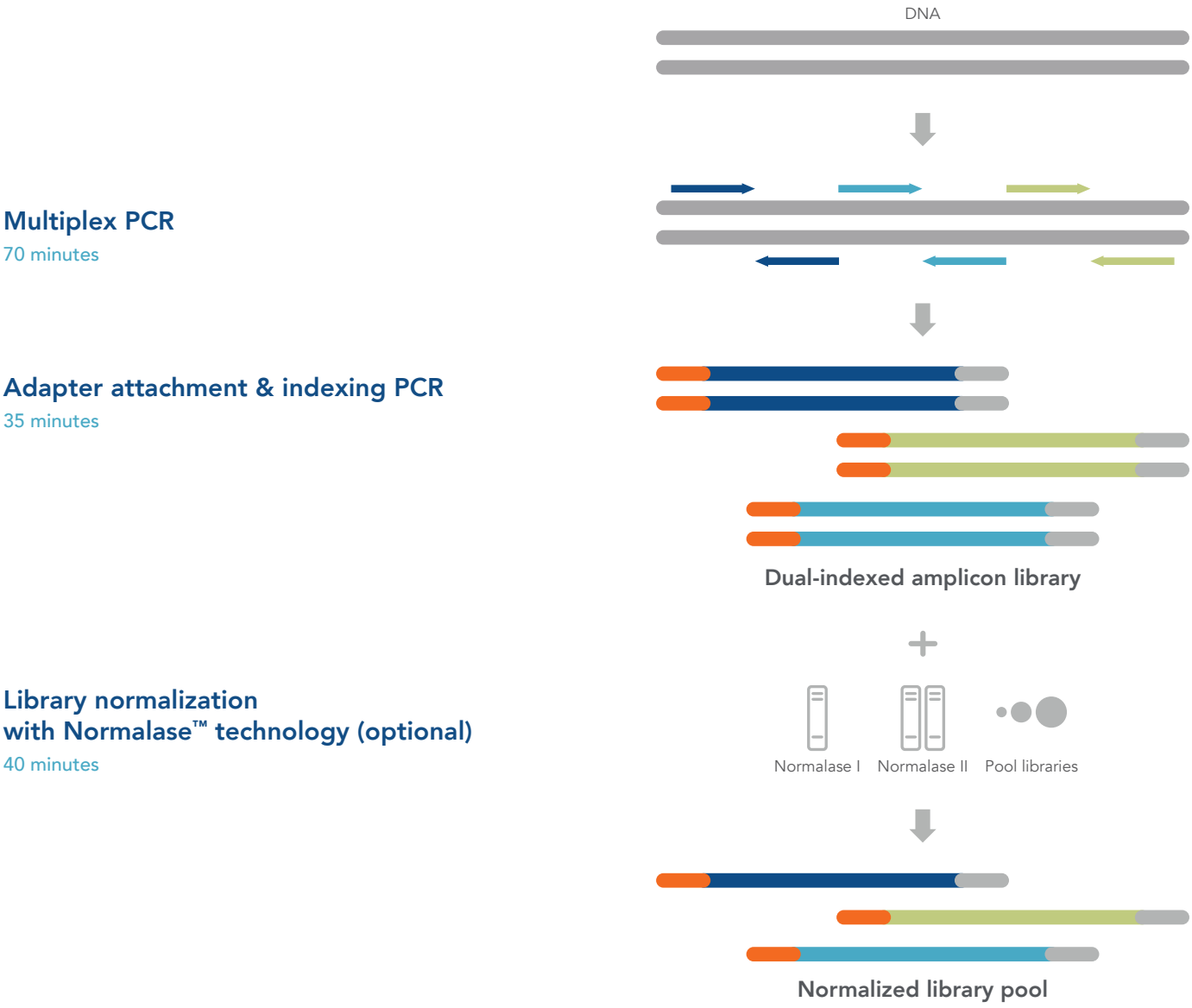

Figure 2. The xGen™ Amplicon Sequencing overview.

Amplicon panel designs can vary significantly. 'Hot spot' panels utilize amplicons that do not overlap and are mostly used for single polynucleotide polymorphism (SNP) genotyping. Panels with overlapping amplicons gather contiguous data from long target sequences (large exons, whole genes, small viral genomes, etc.) in a way that is compatible with short-read sequencing technologies. Many technologies require two tubes for multiplexed PCR when overlapping amplicon coverage is needed to avoid specific primer artifacts. The xGen™ Amplicon Technology uses a single tube workflow for overlapping amplicons without formation of specific primer artifacts, which can reduce sample tracking errors when processing multiple samples while increasing confidence in identification when input material is limiting. This technology can identify variants down to 1% allele frequency by using overlapping amplicons to fully span kilobase range target sequences without any gaps, such as a 30 Kb viral genome.

xGen Amplicon is a one-tube, two-step workflow that takes less than 3 hours to complete. It is integrated with the optional Normalase chemistry for enzymatic library normalization of pools for multiplexed sequencing. The first step performs multiplex PCR, followed by an adapter attachment and indexing PCR step (Figure 2). For samples where an allele frequency of <1% is expected, the xGen™ HS Amplicon Technology uses Unique Molecular Identifiers (UMIs) to identify allele frequencies below 1%.

A subtlety of amplicon sequencing is primer design. Designing functional primers to specifically target the region of interest is often not trivial. New mutations are regularly being discovered and can give important insights. When a mutation occurs under a primer, it can cause a loss of primer binding and function resulting in an intended amplicon not forming; though, both one-tube and two-tube technologies will be able to observe this mutation using the adjacent intended amplicon. In the xGen Amplicon technology, which is a one-tube solution, the super amplicon is formed, thereby allowing observation of the mutation while also avoiding any loss in total genomic coverage regardless of primer dropout. Because other technologies lack the formation of these super amplicons, there is a higher chance of dropouts in genomic coverage which could lead to mutations being missed (Figure 3).

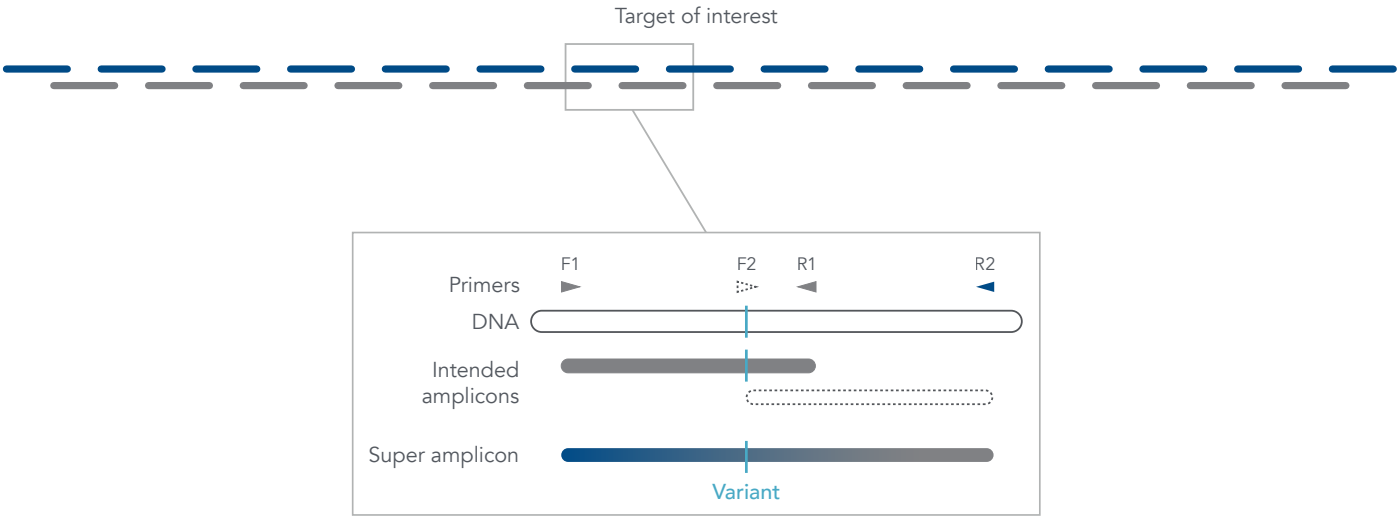

Figure 3: Pinpointing the target of interest. If a variant occurs at a primer binding site, it is still observable in the data. If F2 is no longer functional, the variant will still be seen in the intended (F1–R1) and super amplicon (F1–R2). Coverage between F2 and R2 will come from super amplicons and may be impacted, but will be present. Larger deletions spanning the F2 and R1 binding sites will also be observable from super amplicons, which is only possible with a single tube multiplex PCR. The same situation occurs with all adjacent amplicons (not shown).

CONSIDERATIONS WHEN DECIDING ON A TARGETED SEQUENCING METHOD

Applications

The major factor for determining the sequencing method for your experiment is the research application. Both amplicon and hybridization capture methods can handle challenging samples, including FFPE samples and cell free DNA (cfDNA), and can be used for metagenomic samples or with different species. The amount of damage to a sample can impact data quality and should be considered when planning experiments; for example, oncology research amplicon panels for FFPE and cfDNA are usually designed with a max amplicon size of 150 bp to accommodate DNA crosslinks and short fragments. The read depth requirements are also important to assess. For genotyping/germ line mutations, a lower coverage is needed (50–500X) than for somatic mutations or to observe rare molecules in a sample (>500X–10,000X or greater).

Hybridization capture and amplicon panels can support a wide target size range. As mentioned above, amplicon panels are optimal with smaller target ranges, while hybridization capture panels perform best with larger target ranges. For example, when targeting the whole exome, hybridization capture panels are readily available and produce high on-target rates, but for panels with fewer targets in the Kb size range, amplicon panels demonstrate a higher on-target. Both amplicon panels and hybridization panels can identify SNPs and Illumina™-compatible indels (insertions and deletions). Note that larger panels target more sequences, although require more sequencing with increased cost; keep this in mind when planning your experiments.

Knowing the type of variation and expected allele frequencies is an important part of choosing a targeting technology. Hybridization capture panels can identify larger indels in DNA and both known and unknown RNA fusions when targeting with RNA-Seq libraries. A few amplicon technologies can identify unknown fusions and most, with appropriately designed primers, can identify known fusions. Hybridization capture is also compatible with targeted methylation sequencing from bisulfite converted DNA libraries, whereas amplicon technologies generally are not. Both technologies can report a 1% allele frequency when sequenced to an appropriate read depth. To call lower than a 1% allele frequency or observe ultra-rare molecules, unique molecular identifiers (UMIs) may be required. If using a hybridization approach, the UMIs can be incorporated during the library creation. With an amplicon approach, the UMIs should be incorporated in the first step since amplicon creation is PCR based. xGen HS is an amplicon-based technology that incorporates UMIs for sub 1% allele frequency variant identification.

Total time, throughput, and automation

When thinking of library prep time and the total number of protocol steps, amplicon sequencing presents a significantly shorter workflow. Since the hybridization capture workflow requires libraries are prepared before the target enrichment step, additional time is necessary. Both workflows are easily automatable with scripts readily available.

Hybridization capture scripts are currently available for the following:

- Beckman Coulter Biomek™ i5 and i7 Automated Workstations
- Hamilton NGS STAR™ for Library Prep Workstation
- Perkin Elmer Sciclone™ Workstation

When considering throughput, workflow length is important, but there are other considerations to take into account such as indexing strategy. Most library prep kits, including those that are amplicon-based for DNA, as well as RNA and methylation library prep kits for hybridization capture, offer options for either combinatorial dual indexing (CDIs) or unique dual indexing (UDIs).

SUMMARY

You will undoubtedly look at a variety of factors when choosing sequencing as part of your research investigations. Targeted sequencing can be the most cost-effective choice for many areas of research. Both xGen Hybridization Capture and xGen HS Amplicon Technology offer options that may be uniquely applicable to your lab's goals. Factors, such as target range, workflow, throughput, and automation should be considered when deciding on the right experimental method for your future discovery.

Targeted sequencing handbook

Technical support: applicationsupport@idtdna.com

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**.

For research use only. Not for use in diagnostic procedures. Unless otherwise agreed to in writing, IDT does not intend these products to be used in clinical applications and does not warrant their fitness or suitability for any clinical diagnostic use. Purchaser is solely responsible for all decisions regarding the use of these products and any associated regulatory or legal obligations.

© 2022 Integrated DNA Technologies, Inc. All rights reserved. Trademarks contained herein are the property of Integrated DNA Technologies, Inc. or their respective owners. For specific trademark and licensing information, see www.idtdna.com/trademarks.
Doc ID: RUO22-0863_001 4/22