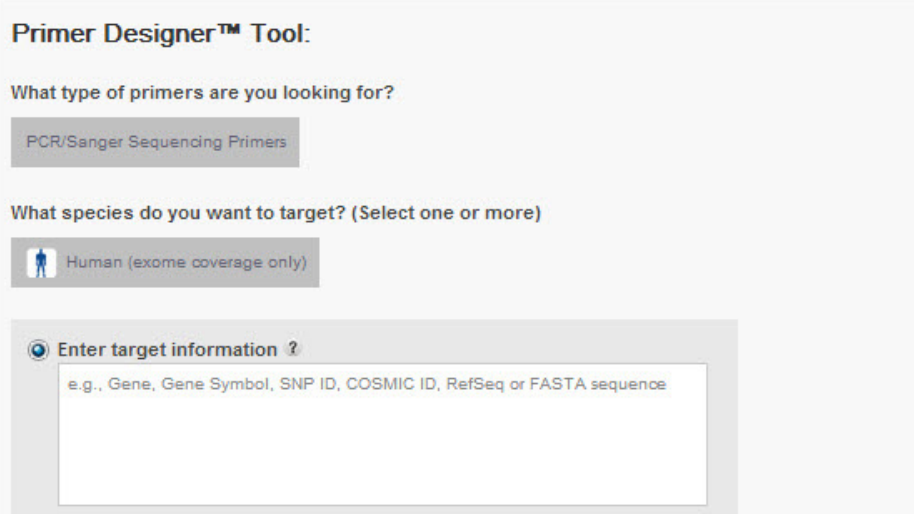
Primer Designer™ Tool

A comprehensive solution for NGS data verification of human exome DNA sequences by optimized Sanger sequencing

Exome or whole genome sequencing by next-generation sequencing (NGS) technologies typically reveals a large number of variations from a reference genome. Many of these variations could underlie variations in phenotypic traits, increased risk of disease susceptibility, or reveal somatic mutations that are potentially oncogenic. Moreover, these variants may occur in genomic regions that are challenging to sequence with high accuracy by NGS or are underrepresented by read numbers. In these situations it is recommended to follow up by verification using an orthogonal method such as traditional fluorescent Sanger sequencing and automated capillary electrophoresis. To facilitate the design process and workflow for NGS data verification of human exome sequences we have developed a complete solution consisting of a comprehensive panel of over 300,000 pre-designed PCR primer pairs covering over 95% of human exons. These PCR primers are designed to be compatible for downstream Sanger sequencing either with



traditional BigDye® Terminator (BDT) sequencing kits v1.1 and v3.1 or the workflow-optimized BigDye® Direct cycle sequencing kit (BDD). The primer designs can be accessed through a user-friendly web portal. The portal also enables direct ordering of custom-synthesized primer pairs. In this application note we demonstrate the utility and convenience of this offering, called Primer Designer™ tool, to the scientific community by showing the process and workflow from NGS-delivered variant call to Sanger-verified result.

Primer Designer™ tool makes primer selection for re-sequencing easy

NGS users may want to verify the finding of a SNP or other variant in the human exome if an actionable decision depends on the accuracy of the result. To that end, the locus of interest has to be re-analyzed with an orthologous method such as PCR-based Sanger sequencing. Designing PCR primers for a locus of interest can be both tedious and challenging. The researcher has to navigate into reference gene sequence files or other DNA sequence resource sites, find the target locus, ascertain that the primers are highly specific for the locus and

life technologies™

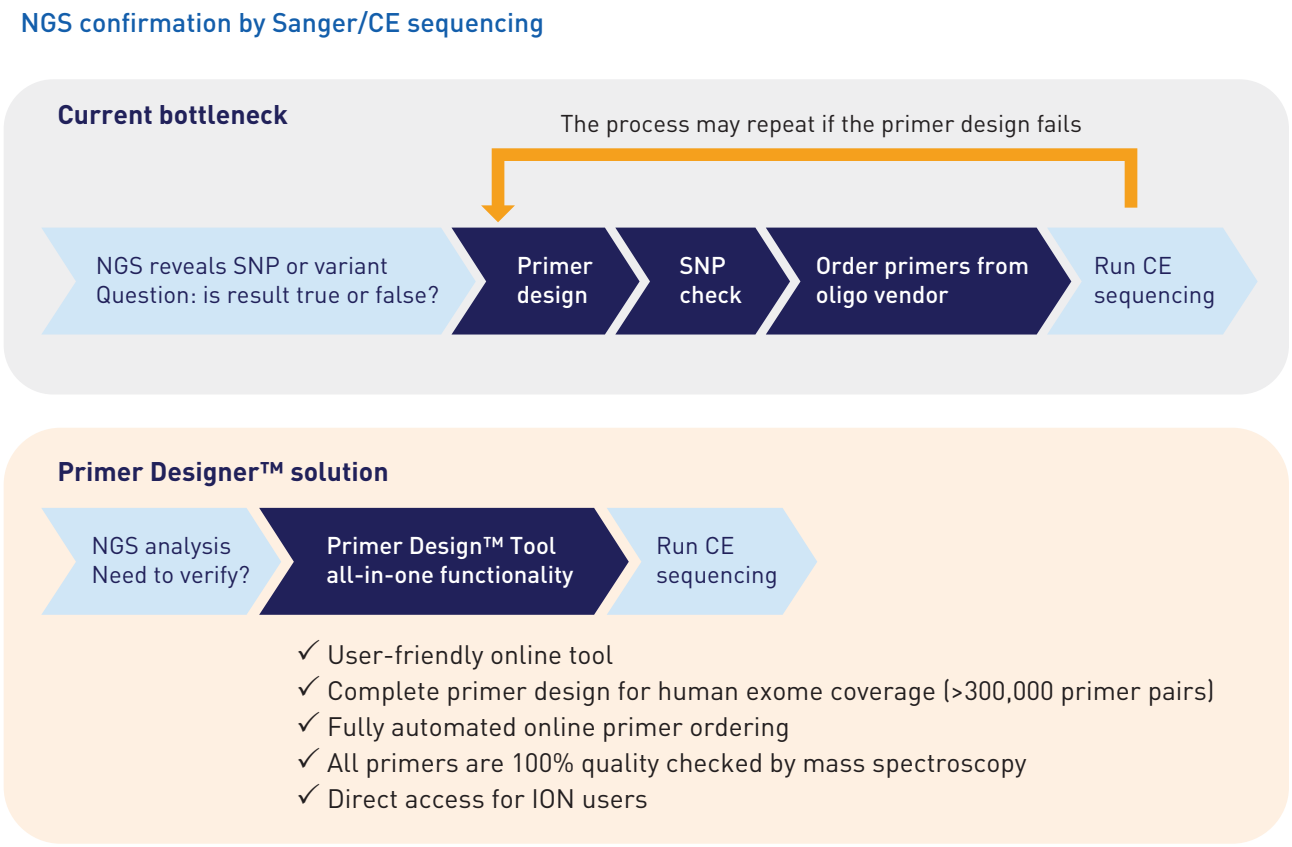


Figure 1: NGS data verification workflow can be tedious and time-consuming (upper pane) or streamlined and convenient using the new Primer Designer™ tool (lower pane).

SNP of interest, and that primer-dimers and undesired secondary structures are avoided. Finally, the design has to be transferred into an ordering sheet of an oligo vendor and then it is a matter of good design and experimental skills to proceed with PCR and cycle sequencing to obtain the DNA sequence from a capillary electrophoresis (CE) trace.

To simplify this process, we have developed the Primer Designer™ tool, which integrates the gene and primer selection, design, ordering and synthesis process with a greater than 95% chance of obtaining a successful and accurate sequencing result in a user-friendly workflow (Figure 1).

The Primer Designer™ tool comprises a virtual design collection of over 300,000 primer pairs that cover 95.9% of human coding exons and 74.9% of human noncoding exons. Complete sequence coverage is achieved for 94% of coding exons and 68% of noncoding target exons. Sophisticated bioinformatics rules and filters were applied to help ensure high target specificity and to avoid SNP locations as well as to minimize the formation of primer-dimers.

Designing the Primer Designer™ tool

The basis for the design collection was the entire human hg19 reference genome (hgdownload.soe.ucsc.edu/goldenPath/hg19/database/refGene.txt.gz). The total number of RefSeq genes was 23,645 and the number of transcripts was 40,344. The total number of exons was 227,706 which added up to a total of approximately 70 million bp. The size range for exons was between 2 and 91,670 bp; intron sizes ranged between 2 and 1,043,912 bp (Figure 2). Great care was taken to avoid common SNPs in the primer designs. Over 11 million SNPs from dbsnp135 (hgdownload.soe.ucsc.edu/goldenPath/hg19/database/snp135Common.txt.gz) were considered.

The typical exon in the human genome can be captured with one amplicon; however, a high number of exons exceeded the intended target length of 500 bp. Since many of these are noncoding, preference was given only to coding exons (Figure 3).

To accommodate long exons into "manageable" amplicons of approximately 500 bp, two or more

amplicons were designed that overlap for 50 bp so that double coverage by forward and reverse sequence was guaranteed (Figure 4).

The Primer Designer™ tool web portal

The Primer Designer™ tool is accessible and free for use on the Life Technologies website (Figure 5): www.lifetechnologies.com/primerdesigner

For the complimentary display of primer sequences a log-in into a user account at Life Technologies.com is required. Although the ordering and purchase of a primer pair as oligonucleotides from this tool are not obligatory, orders are fulfilled promptly if this option is selected. Customers ordering from this tool will benefit from fast turnaround time—in US and Canada desalted primers typically ship within 2 business days, and HPLC-purified primers typically ship within 6 business days from receipt of order—and economical pricing. Each primer order is quality controlled by mass spectroscopy and shipped with a detailed certificate of analysis. Synthesis is done on a 50 nmol scale for HPLC-purified primers, which allows for 100 PCR and sequencing reactions. Desalted primers are synthesized on a 25 nmol scale that allows for many more reactions as the yield is typically 3–5 times higher.

To find the locus of interest the target information field can be queried for a gene symbol like *CFTR* or *KRAS*, a SNP rs identification number, a COSMIC mutation ID, a RefSeq or FASTA sequence file or simply a string of nucleotides (minimum 50).

Alternatively, chromosomal coordinates can be used or an NGS-derived variant call formatted file (.vcf) can be directly uploaded. Most conveniently, users of the Ion PGM™ or Proton™ sequencing platforms benefit from having a direct link from variant to a corresponding primer pair design within the Variant Caller Export Tool in the Ion Torrent™ Reporter™ suite.

Clicking on "Search" will launch the tool and a list of primer pairs will be returned that match the search criteria.

When multiple primer pairs are available it is advisable to use the "View Primer on Map" feature which shows the location of the amplicon generated by the primer pair on a simplified genomic map (Figure 6). By clicking on an exon in the interactive map (depicted as a gray box or bar) the exon number is indicated.

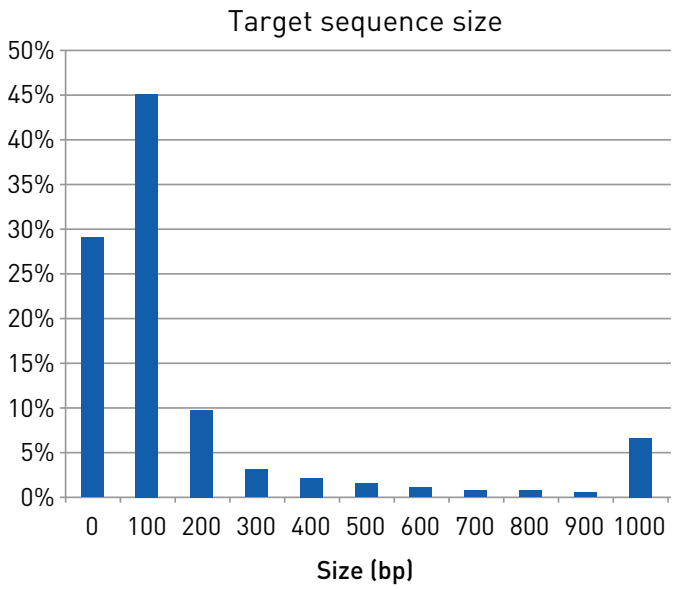


Figure 2: Exon size distribution (in bp) of the hg19 reference genome.

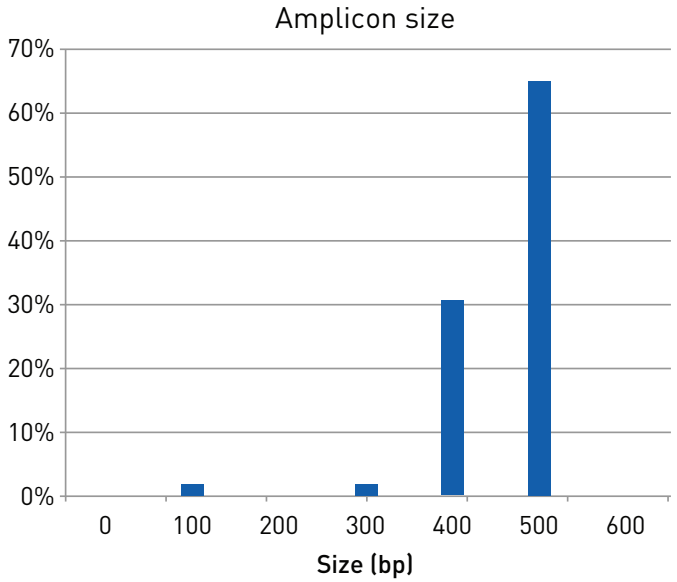


Figure 3: Size distribution (in bp) of the Primer Designer™ human exome collection.

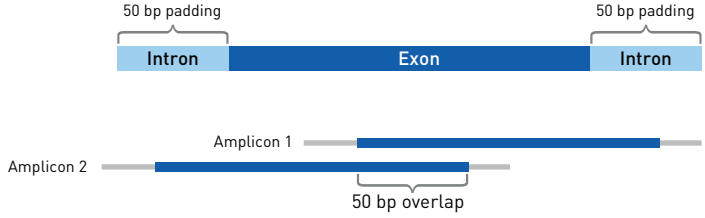


Figure 4: Overlapping amplicons to cover long exons.

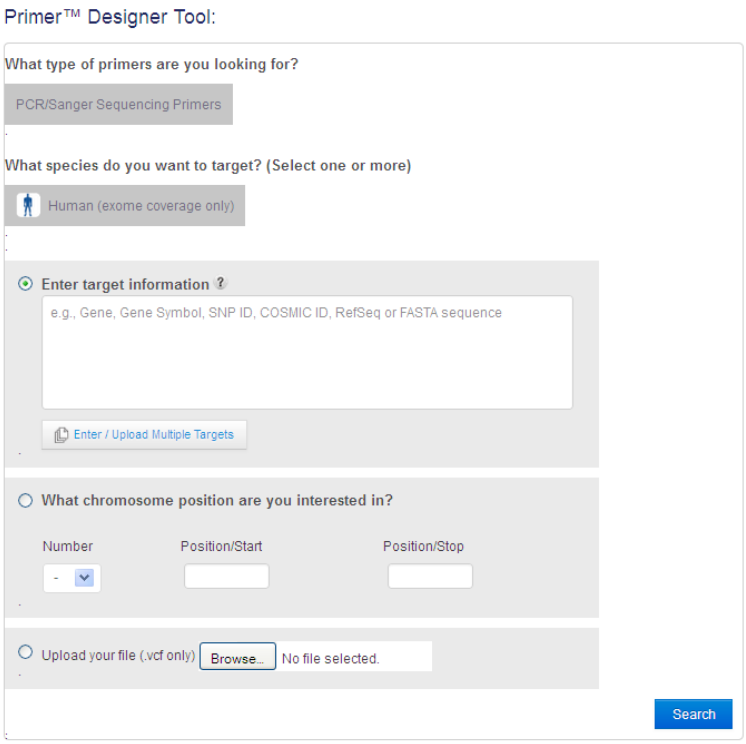


Figure 5: Primer Designer™ tool. The user-friendly interface takes only seconds to find your primers.

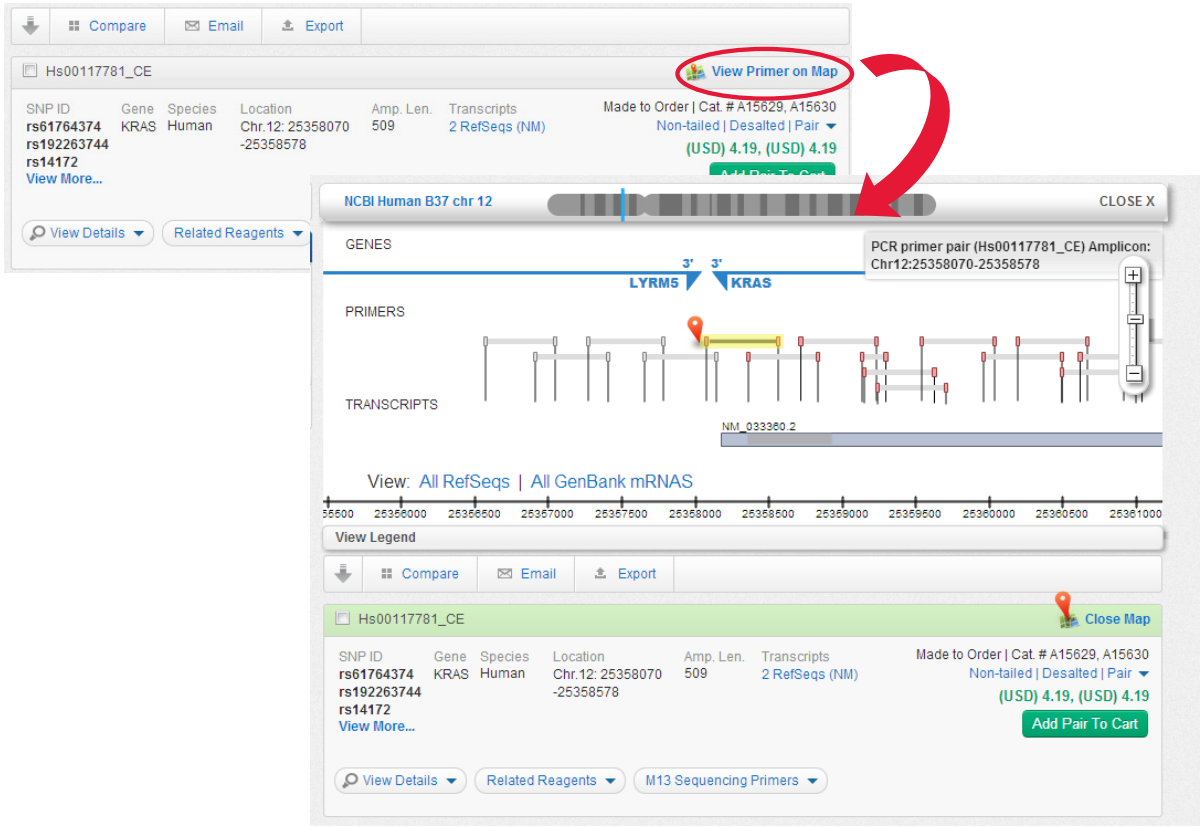


Figure 6: Overview map of primer pairs.

Primer Ordering Options

The Primer Designer™ tool provides the capability for fully-automated online primer order placement from your user account.

The following options are available (Figure 7):

- Primer 5′-tail option: non-tailed vs. M13-tailed
- Purification: desalted vs. HPLC
- Primer pair vs. single primer

Despite the slightly higher cost, we recommend the use of M13-tailed PCR primers since a variant of interest might be located close to either end of a pre-designed assay. With M13-primed sequencing the "blind spot" or PCR-specific priming region of approximately 20 nucleotides at the 5′ end is typically well resolved, leading into the first base of the amplicon target sequence. A visual depiction of this is illustrated in Figure 8. The BigDye® Direct sequencing kit contains specially engineered M13 sequencing primers which exhibit even better 5′ resolution using POP-7™ compared to BigDye®

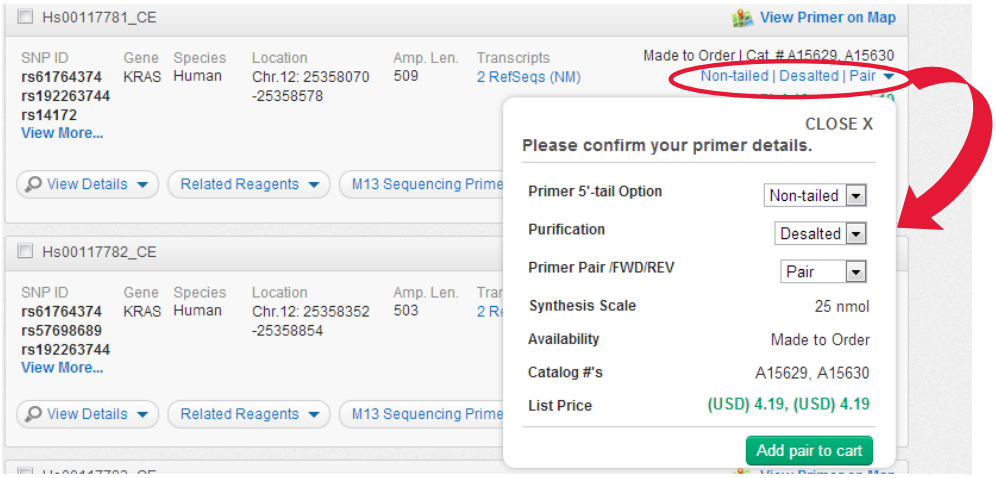


Figure 7: The primer option window.

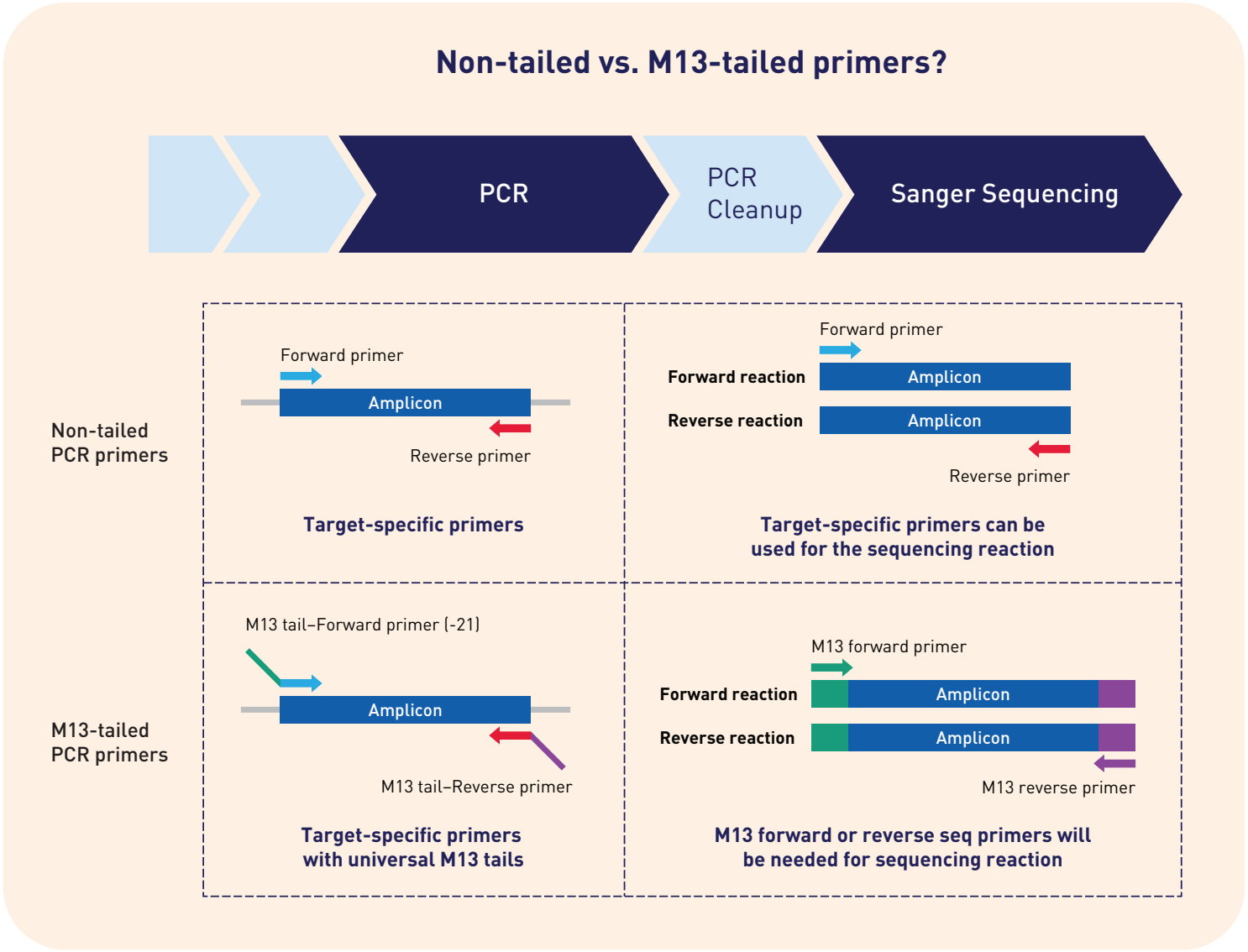
Terminator v3.1 and POP-7™ sequencing. An overview of the sequencing chemistry options are provided in Table 1. We also strongly recommend sequencing both the forward and reverse strands of an amplicon; this is not only good laboratory practice because it validates a sequence by having double coverage from either side, but it may be instrumental to delineate the break points of heterozygous insertion / deletions or strong stops induced by difficult sequences such as extensive homopolymers.

Which BigDye® sequencing kit to use?

The PCR primers designed by the Primer Designer™ tool can be configured to be compatible with any of the existing BigDye® Terminator sequencing kits. The features of these kits are listed in Table 1.

Kit	Benefits	Notes and considerations
BigDye® Direct Cycle Sequencing Kit (BDD)	• Easy-to-use single-plate workflow from PCR to Sanger sequencing—no sample transfer—just add reagents from first to last steps • Kit includes PCR and sequencing reagents priced and formulated ready to go—no need to dilute • Time-to-result (PCR to DNA sequence) as fast as 4 hours for short amplicons (<400 bp) • Superior 5′ resolution on fast POP-7™ polymer (equals more data) • Excellent data quality	• Primers for PCR must be ordered M13-tailed (forward and reverse) in the Primer Designer™ tool—desalted purification is acceptable to yield very good data • Fast time-to-result: quick delivery of desalted primers means data sooner (no waiting for HPLC purification) • HPLC-purified M13-specific primers for sequencing are already included in the kit • Reactions with very high GC content (>70%) DNA require supplemental GC enhancer reagent (Cat. No. 4398848)
BigDye® Terminator v1.1 Sequencing Kit (BDT v1.1)	• Widely-used sequencing reagent • Good 5′ resolution when used in Sanger sequencing with POP-6™ polymer • Non-tailed or M13-tailed primers can be used • Can be used as sequencing reagent and fragment-analysis reagent (SEQ+FA) when used with dye set E5	• 5X dilution buffer is included with kit • Sanger-sequencing run time for POP-6™ is slower than POP-7™ • M13 sequencing primers are not included in the kit but can be ordered HPLC-purified in Primer Designer™ tool; HPLC purification requires extra time
BigDye® Terminator v3.1 Sequencing Kit (BDT v3.1)	• Established standard sequencing reagent • Non-tailed or M13-tailed primers can be used • Uniform high data quality possible for most sequence compositions including homopolymeric sequences	• 5X dilution buffer is included with kit • Resolution at 5′ end—first 25 bases after primer are typically masked or trimmed when run with POP-7™ • M13 sequencing primers are not included in the kit but can be ordered HPLC-purified in Primer Designer™ tool. HPLC purification requires extra time

Table 1: Overview of benefits and considerations for selection of BigDye® sequencing kits.



M13-tailed

- Consistent sequencing conditions with M13 primers
- Single-sequencing mixes (fwd or rev)
- Less pipetting, fewer errors
- More readable data at 5´ end
- Cleaner sequencing traces
- Less baseline noise
- Use with BigDye® Direct or regular BigDye® Terminator kits
- Can be ordered "desalted" for quick turnaround delivery

- M13 tail adds to cost of PCR primer

Non-tailed

- Shorter primers cost less to sythesize
- Same PCR primer is used for PCR and sequencing step

- Not compatible with BigDye® Direct kit
- Same PCR primer is used for PCR and sequencing step—higher risk of failure
- Low resolution at 5´ end; lower base counts
- Sequencing may require optimization
- More pipetting required

Figure 8: Overview of the difference between target-specific non-tailed and M13-tailed primers and their advantages and disadvantages.

For high workflow convenience, short turnaround time-to-result, and overall cost economy we recommend the use of the BigDye® Direct cycle sequencing kit.

How to use

Primers ordered from the Primer Designer™ tool web portal are delivered in dried format. Instructions for how to resuspend and use are described in the Quick Reference guide How to Use Primers Ordered with the Primer Designer™ tool (Pub. No. MAN0008385 Rev. 1.0, downloadable from the Life Technologies website). Standard protocols for PCR, cycle sequencing and purifications can be applied. For the validation study described below the AmpliTaq Gold® 360 Master Mix (Cat. No. 4398876 or similar) was used for PCR and subsequent use with BigDye® Terminator v1.1 or v3.1 sequencing kits. For the BigDye® Direct cycle sequencing kit the PCR reagent contained in the kit was used according to instructions.

Functional validation of Primer Designer™ tool re-sequencing assays

During development of the tool we tested a set of 192 assays (384 primers, Table 2): 24 of these assays were designed based on the finding of 28 SNPs in a human exome sequencing experiment conducted with the Ion Torrent PGM™ platform and 168 additional assays were designed on the top 60 AmpliSeq assays that were ordered from Ion Torrent (www.ampliseq.com).

Four sets of primer configurations were tested with compatible sequencing chemistries using primarily Coriell DNA NA12878 which was also used in the Ion Torrent exome sequencing project.

Oligo set	Oligo design	SEQ chemistry used	# Assays (fwd+rev) tested	Coriell gDNA used
A	Non-tailed/desalted	BigDye® Terminator v1.1	384	NA12878
A	Non-tailed/desalted	BigDye® Terminator v3.1	48	NA12878
B	M13-tailed/desalted	BigDye® Direct	768	NA12878, NA19240
B	M13-tailed/desalted	BigDye® Terminator v1.1	768	NA12878, NA19240
B	M13-tailed/desalted	BigDye® Terminator v3.1	48	NA12878
C	Non-tailed/HPLC	BigDye® Terminator v1.1	96	NA12878
C	Non-tailed/HPLC	BigDye® Terminator v3.1	96	NA12878
D	M13-tailed/HPLC	BigDye® Direct	48	NA12878

Table 2: Set up of Primer Designer™ tool assay validation.

Oligo set	Sequencing success with primer	N tests total	N tests success	% Success
BigDye® Direct	FWD	408	394	96.6%
	REV	408	398	97.5%
	FWD or REV	408	399	97.8%
BigDye® Terminator v1.1	FWD	624	596	95.5%
	REV	624	616	98.7%
	FWD or REV	624	620	99.4%
BigDye® Terminator v3.1	FWD	120	116	96.7%
	REV	120	117	97.5%
	FWD or REV	120	117	97.5%
Total	FWD	1152	1106	96.0%
	REV	1152	1131	98.2%
	FWD or REV	1152	1136	98.6%

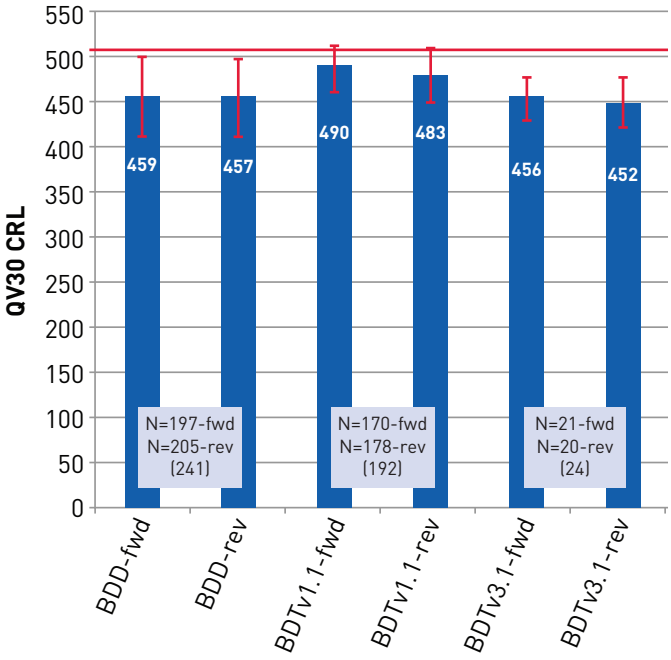
Table 3: Summary of successful PCR to sequence using the three BigDye® sequencing kits.

An important specification of the Primer Designer™ tool was to meet a 95% success rate for on-target PCR product and readable sequence in at least one direction obtained with general sequencing quality metrics of minimal baseline noise (less than 5% peak under peak excluding SNPs and sequence-content-induced baseline noise—both are known limitations of Sanger sequencing) and sufficient signal strength for effective base calling (signal >150 RFU). The data in Table 3 show that this requirement was exceeded with all primer and sequencing kit configurations. Each primer pair design demonstrated amplification of the correct target by at least one chemistry type, indicating there were no manufacturing failures of the oligonucleotides. All failures were related either to primer design (e.g., proximity to difficult sequence content; difficult to amplify under standard conditions)

or human error during reaction setup.

Example of performance metric: contiguous read length (CRL)
Re-sequencing or SNP detection for verification purposes demands superior data quality. A typical metric for high-quality data is the Phred quality value QV30, which indicates a base call accuracy of 99.9% with the probability of an incorrect base call 1 in 1,000. The QV30 read length is calculated on a moving average of 20 bases and reaches the calculated read length for each assay at the point where the 20-bp moving average drops below an average base call of QV30. Because the QV30 metric is a stringent metric of quality it is not uncommon for some 5′ and 3′ trimming to occur giving the appearance of shorter-than-expected amplicons; with the exception of BDTv3.1 / POP-7™ configurations and non-tailed PCR samples, most amplicons read through a majority of the 5′ and 3′ PCR-specific priming regions with good quality base calling. To test the performance of M13-tailed desalted PCR primers we have processed a large panel of assay designs with the three available sequencing chemistries (Figure 9).

Note that the desalted PCR primers were of sufficient purity to generate amplicons which could be readily sequenced to highest the quality using M13 sequencing primers. The purity requirements for PCR primers are less stringent than for sequencing primers. Sequencing primers must not contain n-1 impurities and are therefore typically HPLC purified. The BigDye® Direct cycle sequencing kits contain ready-



% Assays with QV30	82%	85%	87%	91%	87%	83%

Figure 9: Contiguous read length in bp (KB-QV30) with M13-tailed / desalted primers.

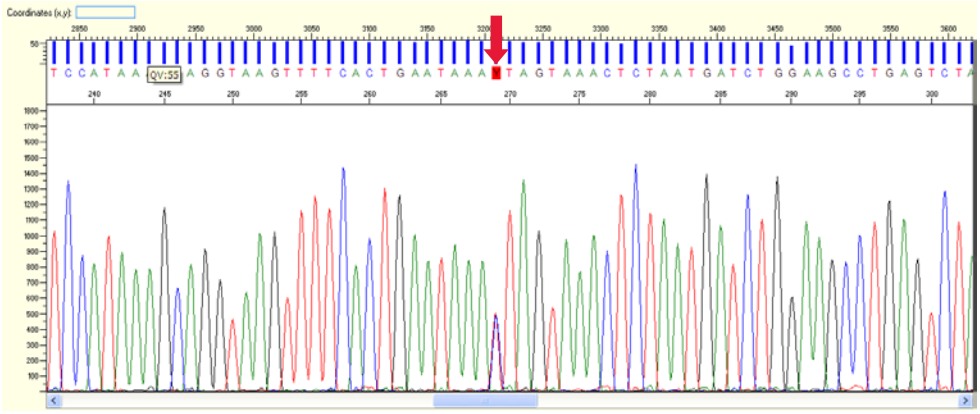


Figure 10: Electropherogram with base calls and quality bars for Hs00257547_CE (exon 10 of MTRR gene): a typical heterozygous SNP (C>T) is detected and called out as a "Y" (pyrimidine) mixed base.

to-use M13 forward and reverse sequencing primers.

Most amplicons (82% and higher) could be sequenced to full length in both directions but in some cases challenging sequence structures can occur which may seemingly hamper analysis but can eventually be resolved to reveal interesting information. An example is shown in Figure 10 with the assay Hs00257547_CE which covers exon 10 of the human MTRR gene.

An intronic SNP (rs2303081) was detected in the center of the sequence data with excellent quality obtained from the read of the reverse strand using the BigDye® Direct cycle sequencing kit (Figure 10, note the "Y" call at position 269). The same sequencing run extended at excellent quality until position 470 when a stretch of 23 homopolymeric "A" nucleotides eventually caused the DNA polymerase to slip and produce unusable data (Figure 11).

Reading from the counter strand a different complication was detected (Figure 12)—early in the sequence at position 45 a sudden deterioration of QV value with appearance of mixed bases occurred. What happened here?

The stretch of mixed bases is due to a heterozygous deletion of 5 bases (del TTTGA) occurring at position 45 in the sequence.

In conclusion, the sequence of this challenging amplicon could be resolved by comparing the forward and reverse reads in combination with our sequence scanner software. We concluded that the amplicon contains a homopolymeric stretch of 23 A followed by a heterozygous indel TCAAA. Furthermore we observed a heterozygous T/C SNP upstream

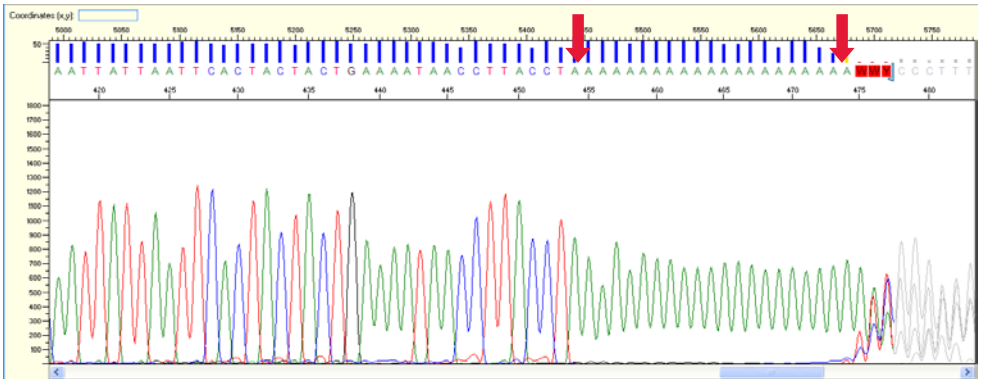


Figure 11: The same assay Hs00257547_CE showing sequences further 3'—a long homopolymer stretch of 23 A bases is detected.

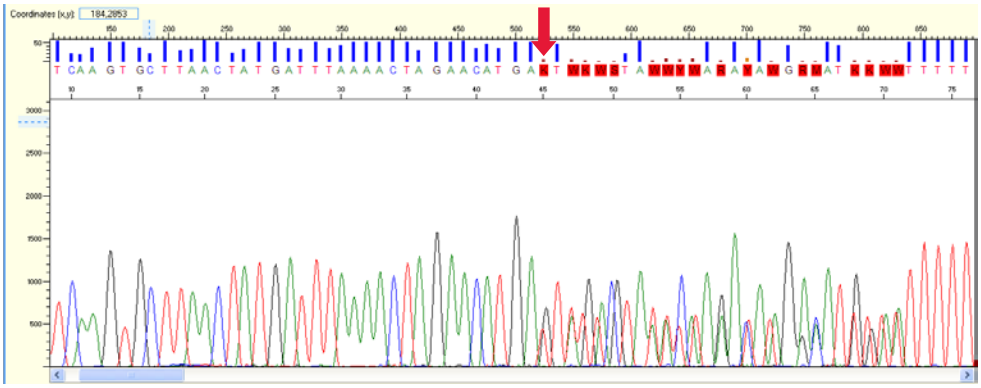


Figure 12: Assay Hs00257547_CE, reading from the counter strand.

Gene	Primer #/Assay	Reference	NGS Genotype	BigDye® Direct CE Genotype	BigDye® Terminator v1.1 CE Genotype	BigDye® Terminator v3.1 CE Genotype	NGS/CE confirmed
NOTCH2	247286	A	A/T	A/T	A/T	A/T	YES
PDE4DIP	247326	C	C/T	C/T	C/T	C/T	YES
PDE4DIP	247326	G	G/A	G/A	G/A	G/A	YES
EXT2	328600	T	T/C	T/C	T/C	T/C	YES
CCND2	115094	C	C/G	C/G	C/G	C/G	YES
ARID2	336029	T	T/C	T/C	T/C	T/C	YES
ALK	192315	A	A/T	A/T	A/T	A/T	YES
MSH6	359011	A	A/G	A/G	A/G	A/G	YES
MSH6	359024	T	C	C	C	C	YES
FN1	205929	T	C/T	C/T	C/T	C/T	YES
BCR	362193	G	G/A	G/A-fwd	G/A-fwd	G/A-fwd	YES
BCR	362194	C	C/T	C/T	C/T	C/T	YES
BCR	362197	G	G/A	G/A	G/A	G/A	YES
MN1	221970	A	A/G	A/A	A/A	A/A	**NO**
CYP2D6	225801	G	G>T	G>T	G>T	G>T	YES
CYP2D6	225801	G	G>A	G>A	G>A	G>A	YES
CYP2D6	225801	T	T>G	T>G	T>G	T>G	YES
CYP2D6	225801	G	G>C	G>C	G>C	G>C	YES
FANCD2	363573	G	G/T	G/T	G/T	G/T	YES
XPC	227180	G	G/A	G/A	G/A	G/A	YES
FOXP1	235541	A	A/T	A/T	A/T	A/T	YES
EPHB1	240070	G	G/A	G/A	G/A	G/A	YES
FBXW7	254918	G	G/A	G/A	G/A	G/A	YES
MTRR	257547	A	A/G	A/G-rev	A/G-rev	A/G-rev	YES
TNFAIP3	281646	C	C/T	C-f/indel-r	C-f/indel-r	C-f/indel-r	YES
TNFAIP3	281648	A	A/C	A/C	A/C	A/C	YES
SYNE1	282846	A	A/G	A/G	A/G	A/G	YES
SYNE1	282961	T	T/C	T/C	T/C	T/C	YES

Table 4: Verification of NGS data. 27 of 28 variants could be confirmed by Sanger sequencing.

of the homopolymer which has been found by others previously and is recorded in dbSNP.

Verification of the NGS Data

Part of the verification data set was a set of 28 SNPs from a human exome sequencing run performed on the Ion PGM™ system using the Ion 318™ chip. Primer Designer™ assays were ordered for all of these SNPs and processed with the three BigDye® sequencing chemistries. Table 4 shows that 27 out of the 28 variants could be confirmed by Sanger sequencing. One potential variant detected by the NGS systems could not be confirmed and was found to be identical to a normal reference sequence.

Conclusions

Primer Designer™ tool is a powerful, versatile and convenient tool for providing researchers in need of urgent NGS data verification with ready-to-use PCR primers that offer high-quality Sanger sequencing data in fast turnaround time. Moreover, any researcher who is interested in re-sequencing or re-analyzing a particular coding segment from the human genome with a different technology will benefit from this comprehensive design collection.

Find out more at **lifetechnologies.com**

life technologies™