PacBio

Application note

Consolidate challenging genes with PureTarget carrier screen panel

Introduction

Carrier screening has demonstrated significant impact in improving human health by determining an individual's carrier status of inherited disease, assessing risk of having affected children, and helping to inform reproductive decision making. Historically, carrier screening was limited to specific conditions or populations, predominantly for cystic fibrosis, hemoglobinopathies like sickle-cell disease and Thalassemias, and ancestry-based screening in Ashkenazi Jewish communities. With the advances in next-generation sequencing (NGS) technologies, this narrow approach has broadened to expanded carrier screening (ECS), determining the carrier status of healthy individuals by screening hundreds of genes at a time. Whereas traditional carrier screening targets specific conditions using legacy technologies like PCR-based techniques, Sanger sequencing, or low-density microarrays, ECS benefits from NGS technology that allow for low-cost, high-throughput screening across a large number of genes. While NGS panels can efficiently screen many genes, genes containing complex variation like structural variants (SVs), copy number variants (CNVs), or tandem repeats still pose difficulties and require ancillary assays (Guha et al. 2024).

Powered by HiFi sequencing, the PacBio® PureTarget™ carrier panel overcomes these challenges with exceptionally accurate long reads that reliably span these variants and simultaneously profile methylation. With a CRISPR-Cas9 system, PureTarget performs PCR-free enrichment without artifact, error, or bias. The PureTarget carrier panel screens for the most challenging variants in 12 genes associated with inherited disease (Table 1). With support for automated library prep and high-throughput sequencing, the PureTarget workflow can help enable time- and cost-efficient screening. By offering more powerful and accurate genotyping in a single streamlined assay, PureTarget helps labs reduce labor and cost while providing more comprehensive characterization of the most challenging genes that impact human health.

This application note details the PureTarget carrier panel targets, presents example performance, and provides a technical overview of the methods, workflow, and analysis pipeline. Together, this presents PureTarget carrier panel as an end-to-end reliable, high-quality assay for carrier screening research with the potential to replace costly ancillary assays.

102-326-653 REV01 SEP2025

PacBio

Gene	Associated disease	Genotyping challenges	Traditional genotyping methods	PureTarget solution
AFF2	Fragile X syndrome, FRAXE type	Tandem repeat expansion, GC-rich sequence, methylation	Triplet-primed PCR	Capture of tandem repeat in ~4 kb fragment and genotyping with TRGT
ARX	Early-infantile epileptic encephalopathy (EIEE1) and Partington syndrome (PRTS)	Tandem repeat expansion	Sanger sequencing	Capture of two tandem repeat loci in ~13.5 kb fragment and genotyping with TRGT
CYP21A2	Congenital adrenal hyperplasia	High homology with pseudogene *CYP21A2P*, copy number variation, gene-pseudogene fusion	Long-range PCR with Sanger sequencing and MLPA	Capture of gene and pseudogene with single pair of gRNAs (~7 kb) and genotyping with paraphase
F8	Hemophilia A	Inversion	Long-range PCR, inverse PCR, PCR-based inverse shifting procedure	Capture of 4 possible inversion breakpoints, mapping with paraphase and read clipping analysis
FMR1	Fragile-X disease (FXS)	Tandem repeat expansion	Triplet-primed PCR, southern blot	Capture of tandem repeat in ~4 kb fragment and genotyping with TRGT
FXN	Friedreich ataxia	Tandem repeat expansion	Triplet-primed PCR	Capture of tandem repeat in ~4 kb fragment and genotyping with TRGT
GBA	Gaucher disease	High homology with pseudogene *GBAP1*, copy number variation, gene-pseudogene fusion	Long-range PCR with Sanger Sequencing	Capture of gene and pseudogene with single pair of gRNAs (~7 kb gene, ~5.5 kb pseudogene) and genotyping with paraphase
HBA1/2	Alpha thalassemia	Large deletion	qPCR, MLPA	Capture of 3 nested regions with 5 gRNAs placed in flanking sequence, genotyping of small deletion with paraphase and large deletion with sawfish
HBB	Sickle cell anemia and beta thalassemia	None	NGS	Capture of ~5 kb region with pair of guides, small variant calling with paraphase
RPGR	X-linked retinitis pigmentosa	Low coverage in ORF15 exon in short read sequencing	Long-range PCR with NGS or Sanger sequencing	Capture of ORF15 exon in ~5 kb region with pair of guides, small variant calling with paraphase
SMN1	Spinal muscular atrophy	High homology with paralog *SMN2*, copy number variation, gene conversion	qPCR, MLPA	Capture of gene and paralog with single pair of gRNAs (~14 kb) and genotyping with paraphase, coverage analysis for copy number of identical haplotype
TNXB	Classical-like Ehlers-Danlos syndrome	High homology with pseudogene *TNXA*, copy number variation, gene-pseudogene fusion	Sequencing and MLPA	Capture of partial exon1 and complete exon2 - exon 13 in *CYP21A2/P* fragment. Variant calling not yet implemented

Table 1. PureTarget carrier panel overview.

PureTarget carrier panel overview

The PureTarget carrier panel enables comprehensive, scalable screening of challenging genes associated with common recessive diseases that are difficult to resolve with traditional sequencing technologies. Compatible with the Revio® system with SPRQ™ chemistry and the Vega™ system, the carrier panel is available in kits that support manual or automated library prep, with the PureTarget kit 24 and PureTarget kit 96, respectively. Up to 96 samples can be multiplexed per SMRT® Cell on Revio + SPRQ, using automated library prep on the Hamilton NGS STAR MOA. Also, 8-48 manually prepared samples per SMRT Cell can be sequenced on Revio + SPRQ or on Vega. A single Revio system can process ~100,000 samples per year, while a Vega system supports ~10,000 samples annually.

The carrier panel includes a set of clinically important but historically "dark" genes that confound short-read sequencing due to tandem repeats, copy number variation, large deletions, or high sequence homology. For example, the *FMR1*, *FXN*, *AFF2*, and *ARX* genes all contain tandem repeat expansions. *SMN1*, *GBA*, and *CYP21A2* each have high sequence homology with a paralog or pseudogene that complicates the ability to determine functional gene copy number and, as a result, carrier status. Lastly, complex structural

variations like large deletions and inversions complicate genotyping of *HBA1* and *HBA2* and *F8*. With PureTarget and HiFi sequencing, genes that are challenging for different reasons and that typically require distinct ancillary technologies to characterize (see Table 1) can be combined. The PureTarget carrier panel also includes *HBB*, *RPGR*, and *TNXB* for convenience and comprehensiveness of the panel.

For clinical research labs it is important to have fit-for-purpose variant calling software. The PureTarget carrier pipeline (PTCP) uses three well-established variant calling tools for HiFi sequencing data: the Tandem Repeat Genotyping Tool (TRGT, Dolzhenko et al. 2024), Paraphase (Chen et al. 2025) for genes with high homology copies, and sawfish for structural variant detection. These tools are stitched together in a WDL-based pipeline along with a QC tool (PTCP-QC) that produces coverage summaries for each sample and target. PTCP is available as a DNAnexus Marketplace App, in the Golden Helix VarSeq software, and available through GitHub for flexible deployment to on-prem servers.

Specification	Metric
DNA input	1–4 µg/sample manual 1–1.5 µg / sample automated
DNA quality[1]	GQN at 30 kb ≥5
Mean target coverage[2]	100-fold or greater per 1 µg DNA per sample
Minimum target coverage	20-fold per sample
Sample multiplexing	96 samples per SMRT Cell on Revio + SPRQ with PureTarget kit 96 8–48 samples per SMRT Cell on Revio + SRPQ and Vega with PureTarget kit 24
Turnaround time (library prep)	8 hours (24-plex manual) 16 hours (96-plex automated on Hamilton NGS STAR MOA)
Methylation	Detected

Table 2. PureTarget carrier panel product specifications.

Performance highlights

Widespread adoption of the PureTarget repeat expansion panel and technical validation of the PureTarget carrier panel demonstrates robust performance in accurately characterizing target variation. Consistent, deep coverage for multiple sample types contributes to this by improving the confidence of results and reducing false negatives. Figure 1 shows typical coverage of the carrier panel for a set of 96 random blood samples sequenced on Revio with SPRQ chemistry. Coverage for the shorter targets (like unexpanded *FXN* and *AFF2*) are 300-fold or higher for Nanobind®-extracted samples. Longer targets like *ARX* (containing two repeat expansions in one target) and *SMN1/2* have coverage greater than ~40X, so genotyping with long and accurate HiFi reads is straightforward. Note the design of *HBA* targets overlapping regions. Samples without large deletions, like the random donor blood shown in Figure 1, do not have large *HBA* deletions, therefore the coverage of these regions is expected to be low.

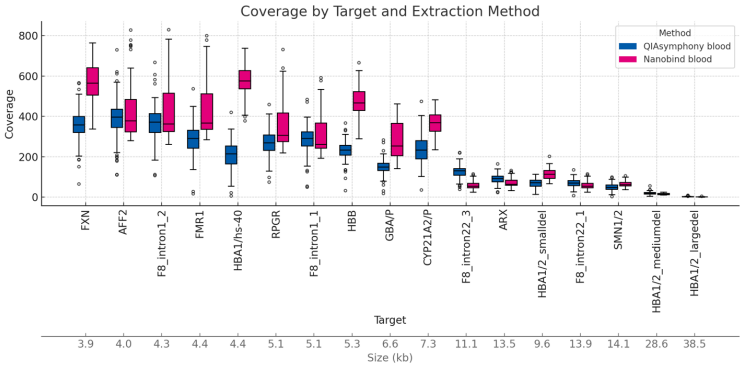

Figure 1. On-target coverage of 96 blood samples processed with the PureTarget carrier panel using automated library prep on Hamilton NGS STAR MOA and sequenced on Revio + SPRQ. Panel targets are sorted by target length and demonstrate that coverage is highest for shorter targets. Two extraction methods are compared: Nanobind and QIAsymphony. Both extraction kits are compatible with PureTarget, but the higher-quality Nanobind samples generally have higher on-target coverage.

1. 50% of mass of DNA molecules longer than 30 kb as measured on Femto Pulse (Agilent).
2. For supported sample types only (Nanobind-extracted blood, lymphoblastoid cells, or saliva – manual PureTarget kit 24 on Revio + SPRQ only, and Coriell lymphoblastoid cell DNA with $GQN_{30kb} \geq 5$). Mean target coverage per 1µg input at max plex (Revio SPRQ 96, Vega 48), higher coverage is possible with higher DNA input and lower plex. Individual target coverage is lower for longer target regions or expanded alleles. The same library loaded on Revio + SPRQ will typically give higher coverage than Vega.

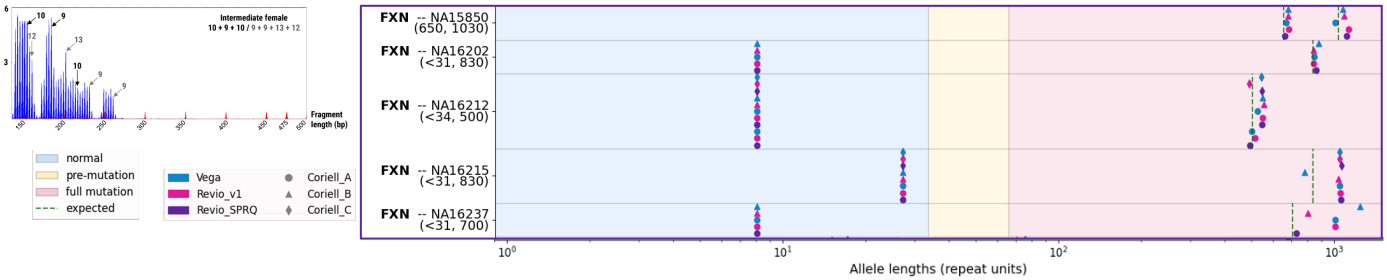

Figure 2. Swimlane plot of five Coriell samples prepared with PureTarget for *FXN*. Each data point represents a technical replicate sequenced on either Revio (pink and purple) systems or the Vega system (blue) with expected size from repeat-primed PCR shown as a dashed vertical line and in text below the sample ID on the left of the swimlane plot. Swimlanes are colored as "normal," "pre-mutation," and "full mutation" depending on repeat length.

For carrier targets with repeat expansions, the PureTarget carrier panel has been technically validated to show consistent repeat sizes across HiFi sequencing systems. This is evident with the sequencing of five reference Coriell samples targeting *FXN* (Figure 2), a GA-rich repeat expansion which causes Friedreich ataxia and is typically genotyped with repeat-primed PCR (Figure 2 inset). When sequenced on both Revio and Vega systems, PureTarget shows highly consistent repeat sizes with TRGT, demonstrating both consistency among technical replicates and agreement with expected size from repeat-primed PCR.

HiFi sequencing on the Revio and Vega systems maintains methylation profiles of native DNA in every run without the need for additional analysis. This enables epigenetic profiling of carrier targets like *FMR1* and *AFF2* where methylation is clinically relevant.

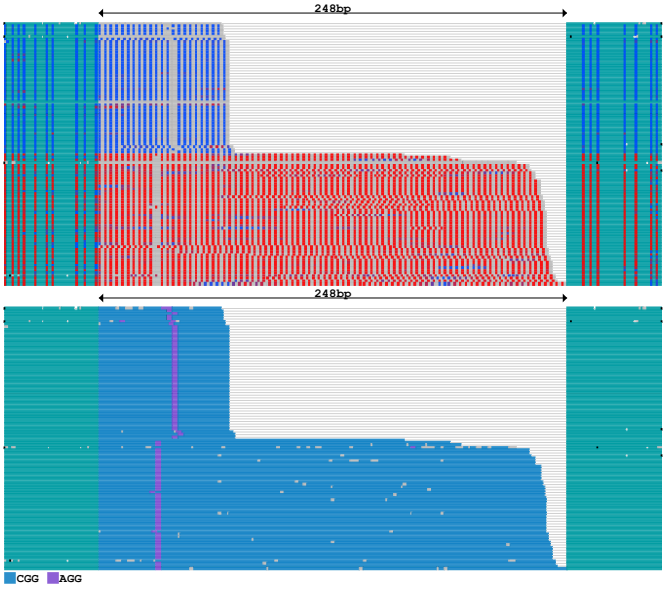

Figure 3. Top panel: waterfall methylation plot of *FMR1* for a carrier female (Coriell sample NA06905). Bottom panel: waterfall allele plot shows AGG interruption motif in purple.

Sequencing of a female carrier (Coriell sample NA06905) prepared with PureTarget shows complete allele-specific 5mC methylation for a pre-mutation length *FMR1* repeat (Figure 3, top panel). Here, 1 µg of DNA was sequenced on the Vega system to validate a CGG repeat expansion based on Southern blot and PCR analysis. The waterfall plot reveals hypermethylation on longer allele 2. The observed genotype from TRGT 2.0.0 is 23 and 79 CGG motifs with an AGG interruption observed on each allele (Figure 3, bottom panel).

Fully phased reads are available for every carrier target with the paraphase analysis tool, enabling clean haplotype examination. An HG03540 reference sample sequenced on the Revio system with SPRQ chemistry was analyzed with paraphase v3.3.1 for *CYP21A2* (Figure 4), which is associated with congenital adrenal hyperplasia. A pair of PureTarget guides captures the full *CYP21A2* gene and its pseudogene. Reads for both the gene and pseudogene are mapped to the gene with paraphase. Five haplotypes are captured showing 2 gene copies (purple and yellow), 2 pseudogene copies (blue and green), and a fusion allele (pink).

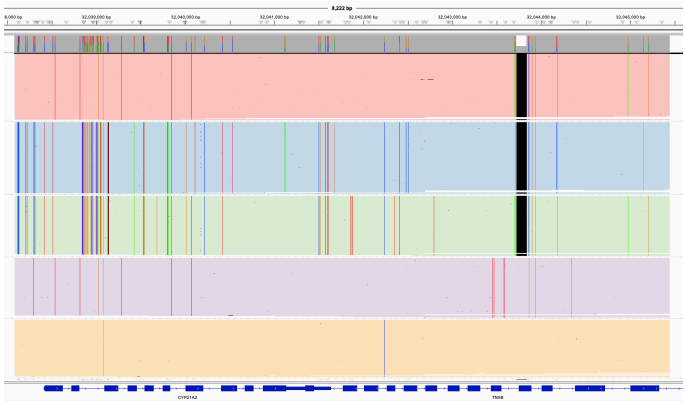

Figure 4. *CYP21A2* haplotypes of HG03540 reference sample prepared with PureTarget carrier panel and sequenced on the Revio system with SPRQ chemistry.

PureTarget method overview

PureTarget uses a single, streamlined workflow that presents several advantages to PCR-based methods or workflows requiring multiple ancillary assays. PureTarget libraries use CRISPR-Cas9 for target enrichment (Tsai et al., 2018, Tsai et al., 2022). The advantages of this approach over other methods include:

- Ability to consolidate genes with tandem repeats, structural variants, and copy number variants into a single assay.
- Retention of methylation signal in the resulting library molecules
- Less size-bias leading to better coverage in repeat expanded alleles
- Coverage in regions of high GC%
- Absence of replication errors and other artifacts

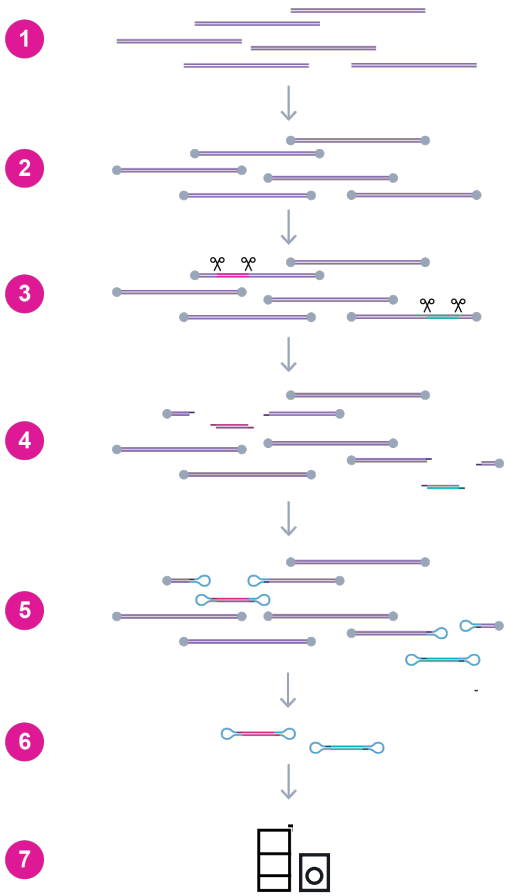

Figure 5. PureTarget enrichment with CRISPR-Cas9.

Figure 5 illustrates the major steps of target enrichment with CRISPR-Cas9. Briefly, high molecular weight (HMW) DNA extracted with Nanobind is dephosphorylated to block the 5' and 3' ends. Next, a complex of Cas9 nuclease enzymes and guide RNAs is used to digest the DNA at precise cut sites upstream and downstream of the repeat regions of interest. Then, barcoded SMRTbell® adapters are ligated following dA tailing of the newly cut ends. Finally, nuclease digestion removes non-SMRTbell molecules. The barcoded SMRTbell libraries can then be pooled before a final cleanup to prepare the sample for annealing, binding, and cleanup and subsequent sequencing.

Supported sample types and DNA extraction methods

To obtain high-quality genomic DNA we recommend Nanobind extraction kits from PacBio. Officially supported sample types are Nanobind-extracted blood, lymphoblastoid cells and saliva (support for manual PureTarget kit 24 on Revio + SPRQ only). Automated Nanobind extraction is available for blood and cells on the KingFisher Apex and Flex systems and the Hamilton NIMBUS Presto. For more details on Nanobind automation capability, refer to the Nanobind HT brochure.

PureTarget has been tested with diverse samples and DNA extraction kits; see Table 3 below for details on common high-throughput blood extractions and Table 4 in the PureTarget repeat expansion application note for a more comprehensive summary of manual and automation extraction methods and non-blood sample types.

If it is not possible to use Nanobind DNA extraction, the QIAsymphony extraction platform and PureGene kits from QIAGEN show good performance with manual and automated PureTarget library prep. Note, the on-target coverage is generally better for QIAsymphony than PureGene (see Figure 1 for an example of QIAsymphony coverage). Other extraction methods in Table 3 like NucleoMag, MagMax, and Chemagic have good results with PureTarget manual prep, but have not been extensively tested with automation or in high-plex.

Sample type	Extraction method	Support category
Blood	Nanobind Whole blood extraction protocol or Nanobind Blood extraction with RBC-lysis protocol	Supported on PureTarget kit 24 and 96
Blood	QIAsymphony (Qiagen)	Demonstrated up to 96-plex and 130 µg total input with PureTarget kit 96
Blood	PureGene (Qiagen)	Demonstrated up to 96-plex and 130 µg total input with PureTarget kit 96
Blood	NucleoMag (Takara)	Demonstrated up to 24-plex and 42 µg with PureTarget kit 24
Blood	MagMax (Thermo Fisher)	Demonstrated up to 32-plex and 40 µg with PureTarget kit 24
Blood	Chemagic (Revvity)	Demonstrated up to 24-plex and 48 µg with PureTarget kit 24

Table 3. Guidance on sample types and extraction methods. To see details on more extraction methods and sample types, refer to Table 4 in the PureTarget repeat expansion application note.

PureTarget analysis software

To sequence PureTarget libraries, select the "PureTarget carrier" application in run design in SMRT® Link v25.3 (also available in SMRT Link Light and SMRT Link Cloud) to enable optimized sequencing conditions for PureTarget libraries (Figure 6). Run design gives users the ability to initiate auto analysis with SMRT Link Target Enrichment once sequencing is complete, which offers quick assessment of sequencing performance such as on- and off-target statistics, target coverage, and sample coverage.

For variant calling, analysis of carrier panel data is consolidated into the PureTarget Carrier Pipeline (PTCP) which provides end-to-end analysis from unmapped BAM files to VCF files and aggregated QC stats. The PureTarget Carrier Pipeline is implemented as a WDL workflow and can be run on local high-performance computing (HPC) systems or in the cloud, specifically via DNAnexus as a Marketplace App or through PacBio compatible partner, Golden Helix, in their VarSeq software. Downstream of PTCP, PacBio Compatible Partners, Golden Helix and Varvis have variant interpretation reporting for PureTarget panels.

Figure 7 shows the major steps of the bioinformatic analysis workflow for the PureTarget carrier panel. Samples are demultiplexed and 5mC methylation probabilities for CpG sites are called on the sequencing instrument. Unmapped BAM files are input into PTCP and the following steps are performed:

- Alignment (pbmm2): Aligns HiFi (and optional fail) reads to the reference genome.
- Tandem repeat genotyping (TRGT): Generates per-sample VCFs containing genotypes for all targeted regions, spanning BAMs of reads used for genotyping, per-locus plots (motif and waterfall), and extracts reads (including optional fail reads) overlapping the specified tandem repeat loci. By default, the following tandem repeats are called in the carrier panel: ***AFF2***, ***ARX*** (EIEE1 and PRTS) ***FMR1***, *and* ***FXN***. Optionally, PTCP can be used to analyze the 38 tandem repeats on the PureTarget repeat expansion panel 2.0, or a custom set of repeat expansion targets.

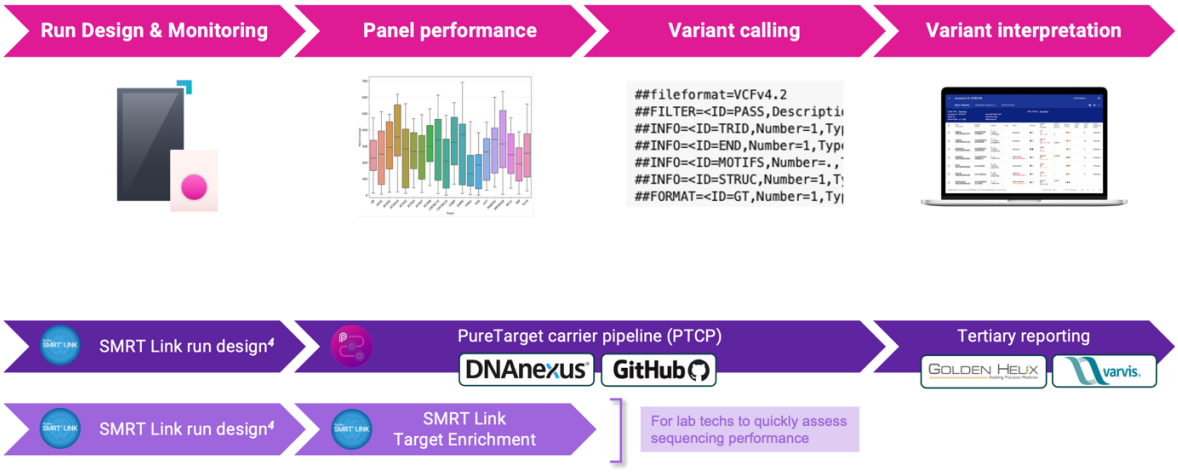

Figure 6. Overview of the different software available for sequencing and analysis of PureTarget carrier screening libraries.

- Gene phasing and analysis (Paraphase): Reads are phased within configured gene families; copy number is estimated, (small) variants are called for each haplotype, and optionally annotated with known variants. Relevant genes for this step include: ***SMN1/2***, ***CYP21A2/P***, ***GBA/GBAP1***, ***F8***, ***HBB***, and ***RPGR***.
- Structural variation calling (sawfish): Aligned reads for configured genes are extracted, realigned, (large) structural variations are called, and reported in per-sample VCFs. Relevant genes for this step include ***HBA1/2***.
- QC reporting (PTCP-QC): Aggregates coverage, mapping quality, and genotyping metrics into both sample-level and cohort-level JSON reports for comprehensive quality control.

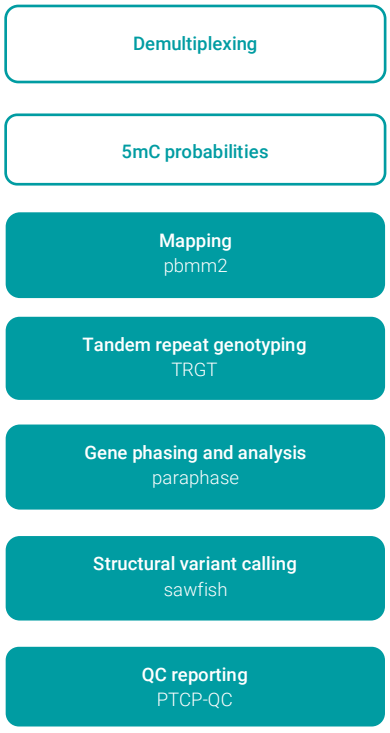

Figure 7. Major analysis steps for PureTarget carrier panel libraries. Analysis in open boxes occurs on the sequencing instrument. Analysis steps in closed boxes are included in the PureTarget carrier pipeline (PTCP).

Conclusion

In summary, the PacBio PureTarget carrier panel offering is an end-to-end workflow featuring high quality DNA extraction with Nanobind kits, amplification-free library prep for a panel of 12 recessive-disease associated genes, variant calling with PureTarget carrier pipeline, and variant interpretation through partners GoldenHelix and Varvis. On Revio with the automated PureTarget kit 96, a lab may process 100,000 samples per year. With Vega, flexible sample batching is possible with 8–48 samples per run.

To learn more about other PureTarget panels, including the ability to create custom panels from the PureTarget catalog of designs or your own genes of interest, refer to the PureTarget custom panel technical note.

References

Chen, X., et al. (2025). Genome-wide profiling of highly similar paralogous genes using HiFi sequencing. *Nature Communications*, *16*(1), 2340.

Dolzhenko, E., English, A., Dashnow, H., et al. (2024). Characterization and visualization of tandem repeats at genome scale. *Nature biotechnology*, *42*(10), 1606-1614.

Saunders, C. T., et al. (2025). Sawfish: Improving long-read structural variant discovery and genotyping with local haplotype modeling. *Bioinformatics*, *41*(4), btaf136.

Resources

Application note – Comprehensive genotyping with the PureTarget repeat expansion panel and HiFi sequencing

Brochure – Nanobind high-throughput HMW DNA extraction

Technical note – A practical guide to amplification-free custom PureTarget panels

102-326-653 REV01 SEP2025