Comprehensive profiling of *CFTR* using PureTarget™ panels, an amplification-free HiFi sequencing assay

Anupam K Chakravarty[1], Jeff Zhou[1], Davy Lee[1], Shreyasee Chakraborty[1], Guilherme De Sena Brandine[1], **William J Rowell[1]**, Heather Ferrao[1], Sarah Kingan[1], Michael A Eberle[1]
[1]PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025

Resolving complex disease genes with native HiFi long reads

PureTarget enables PCR-free investigation of SNVs, repeat expansions, paralogous genes, and structurally complex disease loci.

- Consolidates legacy multi-assay workflows into a single long-read solution
- Captures full-length genes (not just exons) and noncoding variation, preserving long-range phasing
- Retains native methylation and enables multiomics
- Scales across diverse sample and locus types, including SNVs, repeats, paralogs, and structural variants
- Performance across complex loci is further enhanced with SPRQ-Nx chemistry
- Offers off-the-shelf panels for carrier screening (12 genes) and repeat expansions (38 targets), with custom panels also available

PureTarget library prep workflows

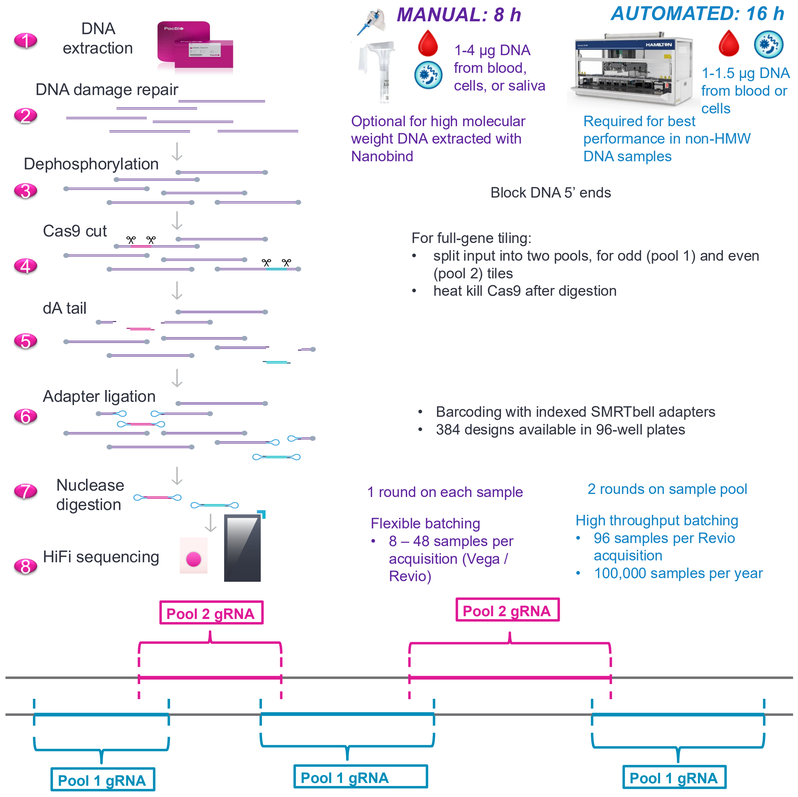

Figure 1. Top: PureTarget library prep. The variant-informed panel uses the standard workflow; the tiling panel requires the split-pool modification at step 4 only. Bottom: Full-gene tiling panel design requires tiled probes in split pools.

Guide RNA design and performance

Iterative gRNA design was performed using population SNP data to maximize enrichment efficiency across diverse ancestries. 67 gRNA designs were tested across the *CFTR* locus.

gRNA design considerations

- Correct strand orientation.
- High on-target activity; low off-target risk.
- Absence of common SNPs in the guide sequence.

Design outcome classes

- **Good**: multiple candidate designs with high on-target, low off-target scores (80% of designs).
- **Limited**: fewer candidates available; acceptable scores after iteration (14%).
- **Sparse**: very few candidates; may require multiple design cycles (3%).
- **Recalcitrant**: no viable designs found; region redesigned or alternative cut sites used (3%).

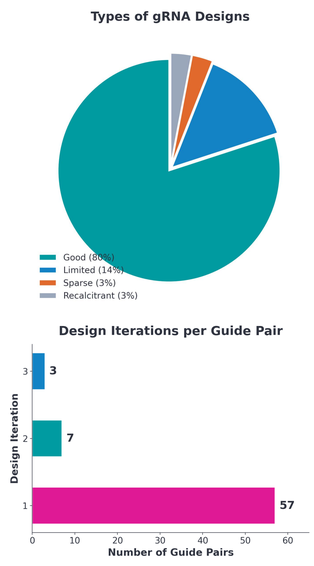

Figure 2. Top: gRNA design outcomes—80% achieved good enrichment (n=67). Bottom: design iterations per guide pair; 85% required only a single design cycle.

Background and motivation

CFTR and cystic fibrosis

The *CFTR* gene spans 189 kb on chromosome 7q31.2 and encodes the cystic fibrosis transmembrane conductance regulator, a chloride channel essential for epithelial fluid homeostasis.

- Cystic fibrosis (CF) affects ~1 in 3,500 newborns; *CFTR*-related disorders (CFTR-RDs) are more prevalent and phenotypically diverse.
- F508del accounts for ~2/3 of CF alleles; >2,000 additional pathogenic variants are distributed across all 27 exons, intron boundaries, and deep intronic regions.
- ACMG 2023 guidance recommends a 100-variant *CFTR* panel, yet current commercial assays capture only a subset of clinically relevant alleles.

Limitations of existing assays

- PCR-based assays are subject to allele dropout and cannot resolve large structural variants.
- Short-read sequencing cannot phase variants or reliably span repetitive regions.
- No existing single assay simultaneously provides variant detection, long-range phasing, and native methylation profiling.

Approach 1: variant-Informed panel

Panel design

- Non-overlapping ~6–10 kb regions flanking all ACMG100 variants and loci prioritized by major diagnostic provider panels.
- Guide RNA selection informed by population SNPs for equitable enrichment across diverse ancestries.
- Single gRNA pool.

Wet lab workflow

- Standard PureTarget library prep protocol, with no modifications to the standard SOP.
- Single Cas9 digestion reaction per sample, compatible with 96-well automation.

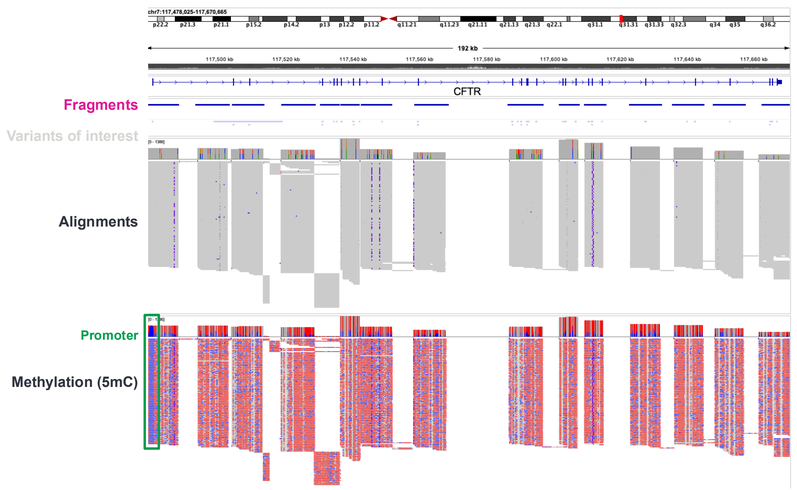

Figure 3. IGV browser view of the variant-informed panel across the *CFTR* locus. Top: gene model, gDNA target file tracks, and variants of interest. Middle: Aligned reads. Bottom: Same dataset with 5mCpG base modification, from blue (low methylation) to red (high methylation). Green box surrounds the *CFTR* promoter.

Advantages of the variant-informed approach

- Simplest path from design to data: standard SOP, no workflow modifications, fastest time to results.
- Streamlined gRNA design targets only the clinically prioritized variant regions.
- Directly compatible with multiplexed carrier screening panels (combine *CFTR* with other genes in one run).
- Excellent coverage and reproducibility at all targeted loci.

Scalability and integration into translational workflows

- Scalable to 96 multiplexed samples per Revio SMRT Cell acquisition.
- Potential throughput of >100,000 carrier screening samples per-year per-instrument.
- Aligns with ACMG 2023 guidance for expanded *CFTR* carrier screening in reproductive medicine.

Approach 2: full-gene tiling panel

Panel design

- 28 gRNA pairs covering the entire 189 kb locus with overlapping ~10 kb fragments organized into odd and even tile pools.
- Overlapping tiles are key: reads from adjacent pools share heterozygous positions, enabling long-range phasing across the entire gene.

Wet lab workflow

- Split Cas9 digestion at step 4 (see Figure 1); all other steps remain identical to the standard SOP.

Analysis pipeline

- DeepVariant (https://github.com/google/deepvariant) + HiPhase (https://github.com/PacificBiosciences/hiphase) for long-range phasing.
- Possible to obtain large extended phase blocks, enabling direct determination of whether variants are in cis or trans.

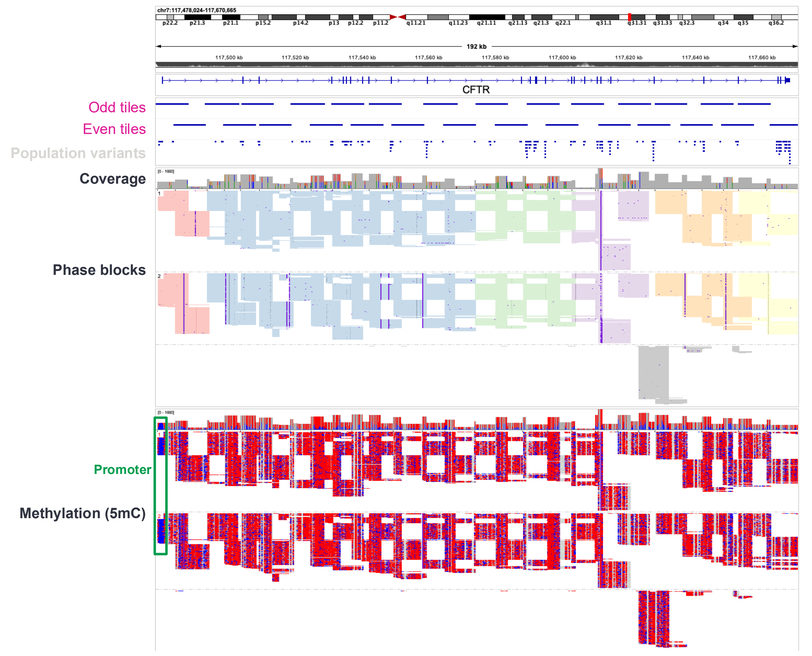

Figure 4. IGV views of *CFTR* full-gene tiling across the ~189 kb locus in one sample. Top: gene model, odd/even gDNA target tile tracks, and common population variants. Middle: Haplotype-phased reads (hap1, hap2, unphased), colored by phase set (PS) tag. Regions with the same color are considered phased together. Bottom: Same dataset with 5mCpG base modification, from blue (low methylation) to red (high methylation). Green box surrounds the *CFTR* promoter.

Advantages of full-gene tiling

- Complete coverage of the entire 189 kb *CFTR* locus, including deep intronic regions inaccessible to targeted panels.
- Long-range phasing can resolve compound heterozygosity directly, without parental samples (*cis* vs. *trans*).
- Multiplexable with other carrier screening genes without loss of efficiency.

Conclusions

1. PureTarget overcomes limitations of PCR- and short-read-based assays, enabling comprehensive *CFTR* interrogation with high accuracy, phasing, and epigenetic context.
2. The variant-informed panel offers a streamlined path to deployment using the standard PureTarget SOP; the full-gene tiling panel adds complete coverage and long-range phasing.
3. Both approaches align with ACMG 2023 carrier screening recommendations, include native methylation profiling, and are compatible with multiplexed workflows.
4. PureTarget provides a rapidly scalable framework with the potential to screen hundreds of thousands of carrier samples per year.

References

1. Tsai YC et al. (2022). Multiplex CRISPR/Cas9-guided No-Amp targeted sequencing. In Genomic Structural Variants in Nervous System Disorders, pp. 95–120.
2. ACMG (2023). Recommendations for carrier screening in reproductive medicine. Genet Med.

Financial Disclosure: Authors are employees and shareholders of PacBio.