PacBio

Application note

HiFi amplicon sequencing for Thalassemia

Introduction

Thalassemia is a group of common hereditary anemias that cause significant morbidity and mortality worldwide, particularly in Southeast Asian, Middle Eastern and Mediterranean populations. The predominant types of thalassemia are α-thalassemia and β-thalassemia, caused by mutations in the *HBA1/2* and *HBB* genes respectively. α-thalassemia occurs in two forms that are associated with significant health problems – hemoglobin (Hb) Bart's hydrops fetalis and hemoglobin H (HbH) disease. Beta thalassemia has three main forms: minor, intermediate, and major, which indicate the severity of the disease (National Organization for Rare Disorders, 2017) A precise diagnosis and prediction of the severity of thalassemia is still challenging, especially when a spectrum of variants and homologous regions are involved. Traditional genetic analysis combines a variety of conventional molecular methods, including reverse dot blot hybridization, Sanger sequencing, GAP-PCR, and Multiple Ligation-dependent Probe Amplification (MLPA) to detect variant types like single nucleotide variants (SNVs), insertion-deletions (indels), and

102-326-551 REV01 OCT2022

PacBio

structural variants (SVs) or copy number variations (CNVs) in *HBB* and *HBA1/2*. (Peng et al., 2022).

Short-read next generation sequencing to analyze thalassemia variants can reduce the need for multiple types of molecular techniques, but challenges remain for rare variants, typically in understudied populations and/or mutations not located in regular regions. Variants in homologous regions of *HBA1* and *HBA2* also remain problematic, often requiring proband (family-trio) analysis.

In this Application Note, we focus on HiFi sequencing of full-length PCR amplicons spanning the 3 key genes *HBA1*, *HBA2* and *HBB*.

PacBio® HiFi reads are long (up to 25 kb) and accurate (99.9%). Targeted HiFi sequencing of amplicons has the ability to span entire genes, allowing for straightforward haplotype construction, detection of structural variants or copy number variants in addition to SNVs and indels. These attributes make targeted HiFi sequencing well suited to analyze both rare and common variants of all types, thus increasing the positive detection rate of thalassemia-related variants.

Your advantages

Advantages over conventional methods have been identified by various research teams, as indicated in Table 1. As described, conventional targeted sequencing can miss rare thalassemia variants. The combination of multiplexed long-range PCR and HiFi sequencing allows the capture of both common and rare variants in *HBA1/2* and *HBB*, which is especially important for more isolated community thalassemia cases, where unexpected, rare mutations may exist, and for areas where ethnic diversity is high and admixtures may be more frequent. PacBio's HiFi data also provides phasing information (for SNVs, indels and other variants), which allows to directly perform linkage analysis.

Workflow overview

Here, we present general instructions for co-amplification of full-length *HBA1*, *HBA2* and *HBB* genes from human genomic DNA (gDNA) samples using barcoded primers in a single round of PCR or a 2-round PCR assay with M13 primers. The recommended input gDNA per sample is 50-100 ng.

Publication title	Authors	Year	Author notes
Analysis of rare thalassemia genetic variants based on third-generation sequencing	Peng et al.	2022	"We enrolled 100 cases that either showed an abnormal hematology phenotype or hemoglobin electrophoresis but had negative conventional genetic diagnosis results. In this study, the CATSA method detected an extra 10 cases of clinically significant variants from *HBA1/2* and *HBB*."
Long-read sequencing on the SMRT platform enables efficient haplotype linkage analysis in preimplantation genetic testing for β-thalassemia	Wu et al.	2022	"The *HBB* gene mutations of the three couples were accurately detected, and the haplotype linked to the pathogenic site was successfully obtained without the need for a proband."

Table 1. Overview of some recent publications and key notes from the authors. (CATSA: Comprehensive Analysis of Thalasemia Alleles)

Multiplexing approach

To increase the samples that can be multiplexed in a SMRT® Cell, primers may be designed with unique barcode sequences and validated with the workflow. PacBio offers 384 barcodes for multiplexing and several barcoding options (amplicons may be barcoded during PCR or during library prep with SMRTbell® barcoded adapters, see Available Materials below).

In the study published by Liang et al. (2021), the authors describe a multiplex long-range PCR strategy for amplifying HBA and HBB gene sequences with 5 primer pairs.

In the study published by Xu et al. (2020), actual primer pairs are described, along with PCR conditions and assay performance.

SMRTbell library construction and sequencing

It is recommended to QC the PCR products prior to SMRTbell library construction (some labs QC the ligation products.) The PacBio procedure for constructing SMRTbell libraries is found here. Finally, the SMRTbell library is prepared for sequencing on PacBio long-read sequencing systems following instructions in the Sample Setup module of SMRT® Link, the PacBio web-based end-to-end software workflow manager.

Recommended analysis tools

HiFi reads may be produced and demultiplexed on PacBio sequencing instruments or in SMRT Analysis off-instrument. HiFi reads for each sample can then be analyzed using pbaa, which generates phased consensus sequences by clustering HiFi reads and enables haplotype phasing for accurate diplotype calls using downstream community tools.

Some recommended tools that have been tested and optimized for HiFi reads are:

- Small variant calling: Google DeepVariant bcftools, and GATK.
- Structural variant calling: pbsv
- Long-read–based phasing: WhatsHap

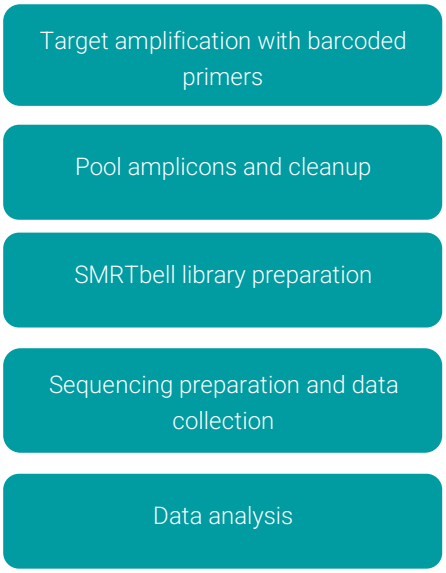

Figure 1. General overview of targeted sequencing workflow on PacBio long-read sequencing platforms with barcoded primers

Customer feedback

"Long read sequencing-based analysis for single gene disorders is revolutionary! It is comprehensive and cost effective. Our lab requires a high degree of sequencing accuracy and HiFi sequencing delivers it!"

Dr Zhou Daixing, Chief Science Office, Berry Genomics

Resources and references

Resources

For more details please visit: https://www.pacb.com/products-and-services/analytical-software/targeted-sequencing/

Available materials

1. Targeted sequencing for amplicons – Best Practices https://www.pacb.com/wp-content/uploads/Application-Brief-Targeted-sequencing-Best-Practices.pdf
2. Preparing multiplexed amplicon libraries using SMRTbell prep kit 3.0 https://www.pacb.com/wp-content/uploads/Procedure-checklist-Preparing-multiplexed-amplicon-libraries-using-SMRTbell-prep-kit-3.0.pdf

References

Liang, Q., et al. (2021) A More Universal Approach to Comprehensive Analysis of Thalassemia Alleles (CATSA). *Journal of Molecular Diagnostics* 239(9), 1195-1204.

Luo, S., et al. (2022) The value of single-molecule real-time technology in the diagnosis of rare thalassemia variants and analysis of phenotype–genotype correlation. *Journal of Human Genetics*, 67, 183–195.

Alpha Thalassemia. NORD (National Organization for Rare Disorders). (2017, January 23). Retrieved September 21, 2022, from https://rarediseases.org/rare-diseases/alpha-thalassemia/

Beta thalassemia. NORD (National Organization for Rare Disorders). (2022, September 12). Retrieved September 21, 2022, from https://rarediseases.org/rare-diseases/thalassemia-major/

Peng, C., et al. (2022) Analysis of rare thalassemia genetic variants based on third-generation sequencing. *Scientific Reports* 12, 9907

Wu, H., et al. (2022) Long-read sequencing on the SMRT platform enables efficient haplotype linkage analysis in preimplantation genetic testing for β-thalassemia. *Journal of Assisted Reproduction and Genetics*, 39, 739–746.

Xu, L., et al. (2020) A New Approach for Identification of Clinically Significant DNA Variants in a-Thalassemia and b-Thalassemia Carriers. *The Journal of Molecular Diagnostics*, 22(8).

Zhong, Z., et al. (2022) A novel 15.8 kb deletion α-thalassemia confirmed by long-read single-molecule real-time sequencing: Hematological phenotypes and molecular characterization. *Clinical Biochemistry*.

102-326-551 REV01 OCT2022