Illumina Sequencing of Gene Targets with HLS-CATCH™

sage science

Application Note: SageHLS

Cas9 endonuclease and custom guide RNAs are used to excise genetic targets from genomic DNA. These are purified and directly sequenced on the Illumina platform.

Introduction

Targeted sequencing typically involves hundreds of hybridization baits or PCR primers that are designed to match protein coding DNA sequences in the human genome. By sequencing the target regions only, valuable genetic information can be collected at a reduced cost. Whole genome sequencing (WGS) provides a more comprehensive analysis of more complex mutations and intronic regions that can inform gene expression. However, genomic phenomena such as structural variation, copy number variation, and pseudogenes are poorly characterized by either approach. Here we present a target sequencing method that can be useful in the analysis of these. The method, HLS-CATCH, uses Cas9 endonuclease to excise intact genomic regions of up to 400kb from cells. Regions are targeted by designing guide RNAs that, with Cas9, can produce a double stranded cut at a single point in the genome. The target regions are purified with the SageHLS system and then directly sequenced.

Methods

Two HLS-CATCH tests were run using a single gel cassette (2 lanes per cassette). Input samples were from the NA12878 cell line and ~275,000 cells (2.5µg equivalent) were loaded into each lane. The Cas9 guide design targets were as follows:

	Gene	chr	Target Size (KB)
Long Panel	BRCA1	17	198
	BRCA2	13	189
	MAPT	17	191
Short Panel	MAPT	17	163
	MUC13	3	103

SageHLS Elution fractions with target fragments were identified by qPCR. Pooled target fractions from one HLS lane (~200-400k target fragments, 10 ng total DNA) were used for library generation using the Agilent SureSelectXT Low Input Target Enrichment System for Illumina Paired-End Multiplexed Sequencing Library kit (using manufacturer protocol vs B0, using 11 cycles of pre-enrichment PCR) -- without SureSelectXT enrichment and secondary amplification steps.

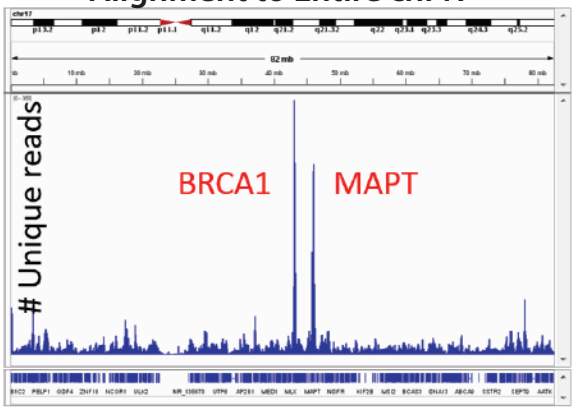
Collaborators gratefully acknowledged:
Melissa Smith, Ethan Ellis, James Powell, Ayesha Rasool, Maya Stahl, and Robert Sebra

Icahn Institute and Dept. of Genetics & Genomic Sciences
Icahn School of Medicine at Mount Sinai

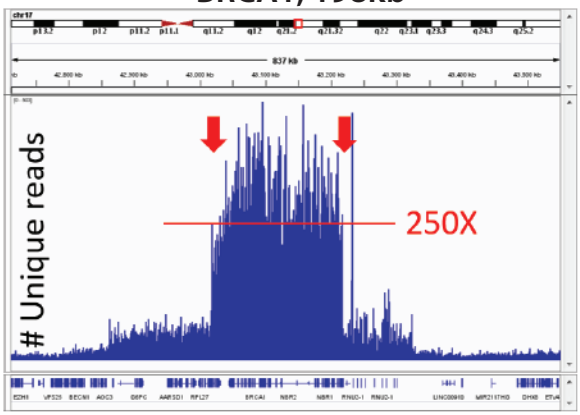
The two libraries were pooled and sequenced on one lane of a HiSeq2500 in Rapid Run mode. Output was aligned to hg38 with BWA-MEM, and PCR duplicates were removed using Samtools. Coverage data was calculated using Bedtools, and displayed in IGV. After removal of duplicates, the Long panel sample produced 186m reads (unique, mapped, pe), and the Short panel generated 358m reads.

Results - Long Panel

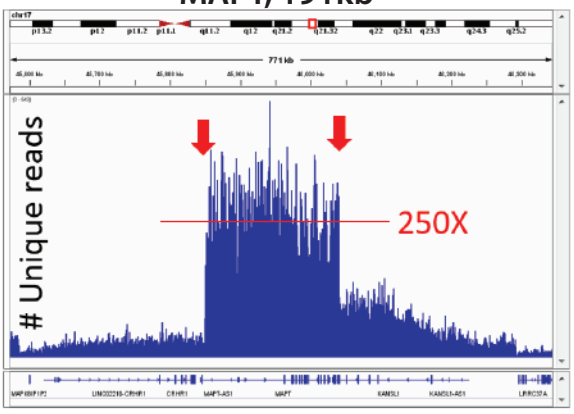
Alignment to Entire chr17



BRCA1, 198kb



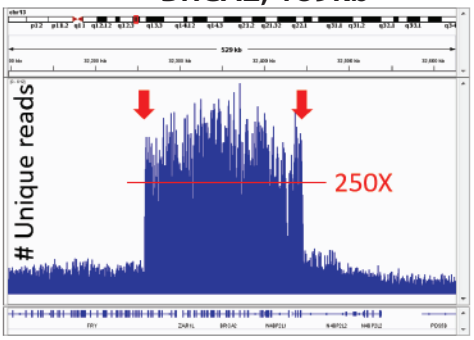
MAPT, 191kb



* Arrows indicate positions of guide RNAs.

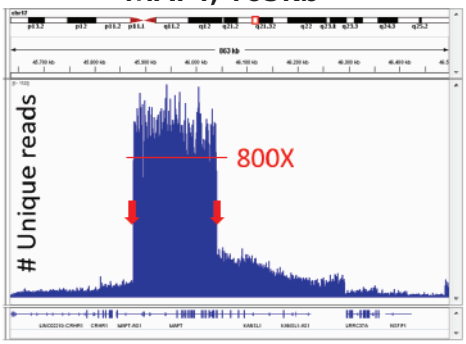
Results - Long Panel (cont'd)

BRCA2, 189kb

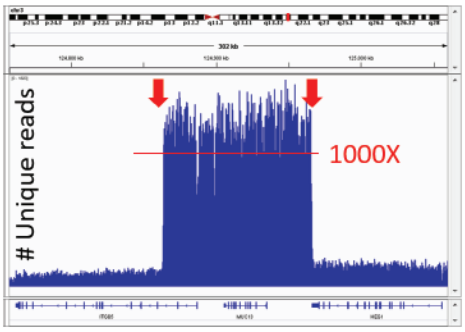


Results - Short Panel

MAPT, 163kb



MUC13, 103kb



Summary

HLS-CATCH can provide more comprehensive coverage when compared to other target enrichment methods. Intronic and flanking sequence around the gene can be included as targets, thus providing analysis of intronic regions such as splice junctions or promoters. The flexibility of design of guide RNAs can also be advantageous as the target regions can easily adjusted with the design of new guideRNA pairs.

A depth of coverage between 250 to 1000X was obtained when the target sequences were aligned. This could provide a very economical alternative to WGS for studies that are focused on complex genes or loci. Using targeted endonucleases offer another advantage over target enrichment and WGS in that assembly and alignment are conducted against a defined region (by size and sequence) rather that an entire genomic reference. This reduces mismatches or ambiguities, particularly for difficult analyses such as pseudogenes.

HLS-CATCH can be multiplexed, and up to 20 guide pairs per run have successfully been used in other studies. An absolute limit to multiplexing has not yet established.