illumina®

Amplicon Sequencing from FFPE tissues on the MiSeq® System

Simple and fast workflow for accurate detection of rare variants even from degraded DNA.

Introduction

The MiSeq personal sequencing system delivers the fastest time to answer with the easiest workflow available. Using Illumina's proven TruSeq® sequencing by synthesis (SBS) reversible terminator chemistry, greater than 1 Gb of data can be achieved in a single paired-end sequencing run. This application note describes sequencing indexed amplicons of KRAS oncogene exon 2 region from genomic DNA extracted from formalin-fixed, paraffin embedded (FFPE) paired tumor/normal samples. From a single MiSeq run, coverages in excess of 150,000-fold were observed for each amplicon, sufficient to detect variants close to the limit of detection (~ 0.5%), even from highly degraded DNA.

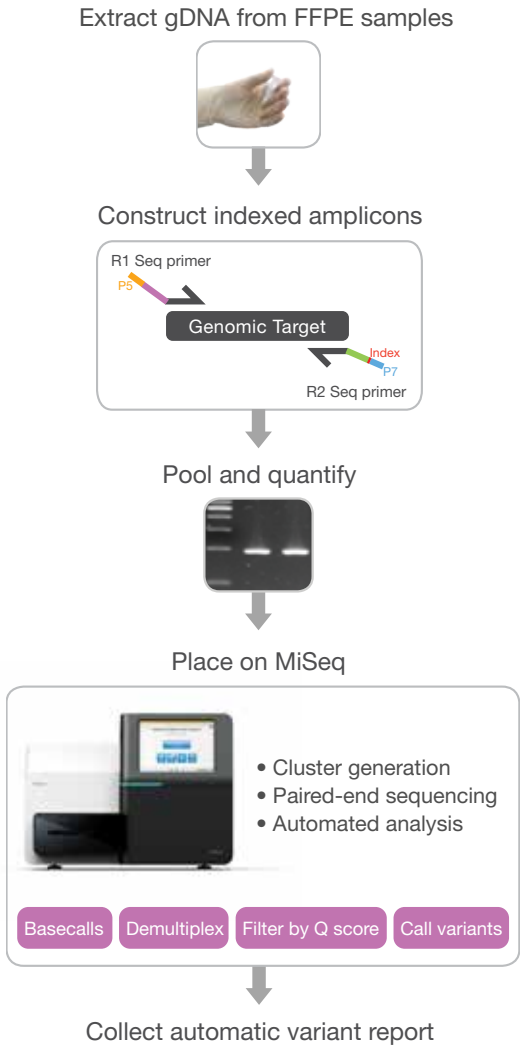
Amplicon Sequencing Workflow

Genomic DNA (gDNA) was extracted from five samples, including NA18507 control DNA (Corriell) and two matched tumor/normal samples for gastric and rectal FFPE tissue from Proteogenex. Using a one-step PCR protocol, 76-bp amplicons corresponding to sections of KRAS exon 2 were constructed using identical SBS-compatible forward primers and indexed reverse primers for each sample. After PCR, amplicons were pooled and quantified using a Bioanalyzer. The pooled amplicons were loaded into the reagent cartridge and placed into the MiSeq instrument along with the flow cell. All steps from amplicon generation, through cluster generation, 2 × 77 bp paired-end sequencing run, and data analysis were performed on the MiSeq in 15 hours (Figure 1). MiSeq on-board analysis generated a variant report on observed sequence compared to a reference sequence. In this data set, the detection threshold was conservatively set at 1.0%.

Results and Data Analysis

The primary run metrics, including total number of reads, mean reads per sample, and total basecalls above Q30 are shown in Figure 2A. The average amplicon coverage was 170,963×. Analysis of the non-reference calls in the NA18507 control indicates that the raw read accuracy was 99.8%, which was confirmed in the FFPE-extracted samples. A single C to G nucleotide variant was observed in the rectal tumor sample in 27.5% of reads (Figures 2B and 2C). This variant was also observed at a much lower level (1.1%) in the adjacent 'normal' rectal tissue. The presence of the same variant in a sample taken adjacent to the tumor indicates that the normal sample contains a detectable percentage of cells containing the variant. No variants were identified in gastric tumor or normal tissue, or in the control DNA.

Figure 1: Amplicon Sequencing Workflow



Amplicons are prepared directly from gDNA by a single-step PCR that adds sequencing primers and index sequences, pooled, quantified, and loaded onto the MiSeq instrument. Cluster generation, paired-end sequencing, and data analysis are performed directly on the instrument, requiring no user intervention. The data analysis workflow includes base calling, demultiplexing, quality filtering, and variant calling on the instrument computer, automatically generating the variant report.

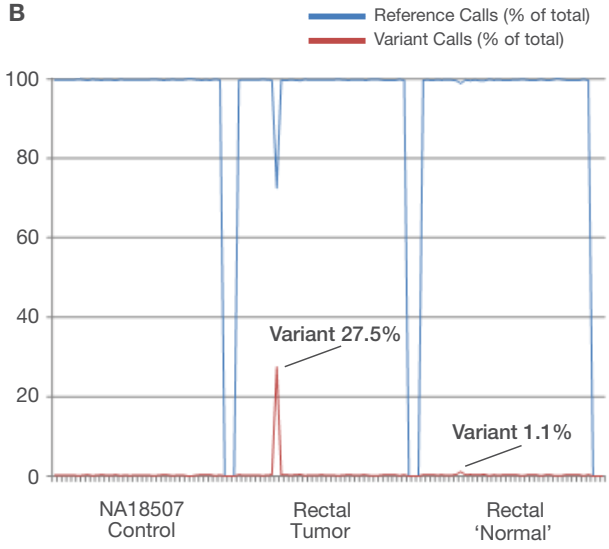
AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
TCAACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTT
CGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACC
AACGTACCATTAAGAGCTACCGTGCAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAA
AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
GATTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAACGAGC
CGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAACAGTA

Figure 2: KRAS Exon 2 C->G Mutation in Rectal Tumor and Normal FFPE Samples

A

Number of reads (1 & 2)	2.11 M
Mean reads per sample (index)	352,177
Total basecalls > Q30	94%

B



C

Sample	Coverage	Noise (%)	Variants
NA18507 Control	179,630	0.21	None
Rectal Tumor	177,392	0.24	27.5% C→G
Rectal Normal	178,667	0.27	1.1% C→G
Gastric Tumor	176,530	0.23	None
Gastric Normal	161,866	0.28	None
Average	170,963	0.25	

A. Primary run metrics of indexed amplicon sequencing.
B. Percentage of variant and reference calls in control, rectal tumor, and rectal normal samples.
C. Coverage, baseline noise and variants for all samples.

Conclusions

Preparing amplicons for sequencing on the MiSeq instrument is very straightforward, requiring a single-step PCR to amplify and add tailed adapters and indexes directly from genomic DNA. Indexed samples are ready for sequencing in three hours with no complicated enrichment steps, beads, or additional instrumentation required. Once samples are loaded into the reagent cartridge, the MiSeq instrument performs automatic cluster generation, paired-end sequencing, and data analysis without user intervention.

In this study, five amplicons were multiplexed and sequenced at very deep coverage. A sequencing run on the MiSeq instrument has the capacity to generate up to 5 million clusters to accommodate up to 48 indexed amplicons from 48 samples simultaneously, yielding an average of 2,500× coverage. The KRAS amplicons in this study were 76 bp in length. The MiSeq platform can produce overlapping paired-end reads to create fused reads covering amplicons of up to 300 bp. However, due to the highly degraded nature of FFPE DNA, Illumina recommends amplicons of less than 200 bp from FFPE sources.

Sequencing results on the MiSeq instrument were extremely accurate, due to filtering out data with predicted accuracy of less than Q30. For this data set, the raw read accuracy was 99.8% with very low average background noise (< 0.25%). The amplicon sequencing results clearly show robust detection accuracy of low abundance variants, which is a critical requirement for cancer mutation studies. In addition, the ability to accurately sequence amplicon DNA from FFPE samples opens up an enormous repository of archival information for studying the genetic changes underlying cancer development and progression.

Learn More

Visit www.illumina.com/miseq to learn more about the next revolution in personal sequencing.

Illumina • 1.800.809.4566 toll-free (U.S.) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

FOR RESEARCH USE ONLY

© 2011-2012 Illumina, Inc. All rights reserved.
Illumina, illumina*Dx*, BaseSpace, BeadArray, BeadXpress, cBot, CSPro, DASL, DesignStudio, Eco, GAIIx, Genetic Energy, Genome Analyzer, GenomeStudio, GoldenGate, HiScan, HiSeq, Infinium, iSelect, MiSeq, Nextera, Sentrix, SeqMonitor, Solexa, TruSeq, VeraCode, the pumpkin orange color, and the Genetic Energy streaming bases design are trademarks or registered trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owners.
Pub. No. 770-2011-008 Current as of 06 February 2012

illumina®