A novel technique that distinguishes low-level somatic DNA variants from FFPE-induced artifacts in solid tumors by next-generation sequencing (NGS)

Nitin Udar1, Robert Haigis2, Thomas Gros1, Nick Kerry3, Bret Barnes1, Dmitry Pokholok1, 1 - Illumina, 5200 Illumina Way, San Diego, CA; 2 - Institute of Personalized Medicine, University of Alabama at Birmingham, Birmingham, AL; 3 - Broad Institute and Massachusetts General Hospital, Boston, MA

Abstract

Several applications, including somatic mutation detection in cancers, require methods that can efficiently detect low-frequency variants (< 1% minor allele frequency, MAF) in DNA extracted from formalin-fixed-paraffin-embedded (FFPE) tissue. We have developed a novel technique that can distinguish true variants from fixation artifacts with high sensitivity and specificity by analyzing each of the two DNA strands independently. "DualStrand" Custom Amplicon Technology was used to generate sequencing libraries and deep sequencing was carried out to an average depth of 20,000X with a minimum of 1000X. The targeted re-sequencing assay* investigates ~14 kb of exons in 26 genes commonly mutated in solid tumors. Testing of more than 200 samples with a MAF ≤5% showed that the presence of a large number of potentially false positive calls when data from only one strand of DNA was analyzed, but the number was significantly reduced (e.g. ~30% for p53) when both strands were considered. Conclusion: This technique can distinguish FFPE artifacts from true variants and therefore provides increased accuracy for the detection of low-frequency variants by NGS. *Research Use Only

Materials and Methods

DNA was extracted from 44 FFPE samples from multiple tissue types using the Qiagen QIAamp Kit. A qPCR-based assay was used to qualify the extracted material for further processing. Genes of interest were targeted using an extension and ligation based amplicon library prep assay specific for each of the two strands of DNA. Index sequence was incorporated into the tagged primers to identify the amplicon and sample. The resulting two independent libraries were combined and sequenced on the Illumina MiSeq instrument. Average depth of coverage was 20,000X. A minimum read depth of at least 1000X coverage for obtained for all amplicons. The reads were aligned and demultiplexed and variant frequencies calculated using information from both strands.

Process Overview

DNA Extraction

Sample QC

Library Preparation

MiSeq® Sequencing

Data Analysis

Sample Qualification

DNA was extracted from 44 FFPE samples from multiple tissue types using the Qiagen QIAamp Kit. A qPCR-based assay was used to qualify the extracted material for further processing. Genes of interest were targeted using an extension and ligation based amplicon library prep assay specific for each of the two strands of DNA. Index sequence was incorporated into the tagged primers to identify the amplicon and sample. The resulting two independent libraries were combined and sequenced on the Illumina MiSeq instrument. Average depth of coverage was 20,000X. A minimum read depth of at least 1000X coverage for obtained for all amplicons. The reads were aligned and demultiplexed and variant frequencies calculated using information from both strands.

Process workflow:
1) Sample QC to assess DNA quality
2) A multiplex assay that involves deep sequencing using next generation sequencing of targets on 26 genes focused on regions implicated in lung, gastric, colon, melanoma and ovarian cancers.
3) A bioinformatics pipeline that can distinguish true variants from fixation artifacts.

Sample Qualification

Library success was best predicted by DNA amplifiability measured by comparison to (ΔΔCt) non-FFPE reference DNA (QCP). We achieved a 100% library success rate with ΔΔCt <1 and qPCR profile.

Results

More than 200 samples were extracted and assayed using deep sequencing. Deep sequencing at 1000X minimum coverage for all amplicons allowed the detection of low frequency variations >5% MAF. The two strand approach, increased accuracy due to the independent confirmation of variants on both strands of DNA in a single assay and therefore distinguished the artifact events from true variations. Minimum coverage of 1000X for all amplicons means optimal utilization of content. Unlike genotyping assays, this sequencing assay can report a variety of variations including SNPs, as well as indels. In addition sequencing reports variant or wild type calls at all positions in the assay for a more comprehensive view of variation.

Conclusions

This novel technique for detecting low frequency somatic variation in FFPE samples by next targeted generation sequencing is able to differentiate true variation from fixation artifacts producing high analytical sensitivity and specificity (>99.9%) and a low limit of detection (down to 3% variant frequency).

© 2013 Illumina, Inc. All rights reserved. Illumina, IlluminaDx, BeadArray, BeadXpress, cBot, CSPro, DASL, Eco, Genetic Energy, GAIIx, Genome Analyzer, GenomeStudio, GoldenGate, HiScan, HiSeq, Infinium, iSelect, MiSeq, Nextera, Sentrix, 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.