ABSTRACT

Many human diseases are associated with genetic polymorphisms. Resequencing candidate regions can provide valuable information about the genetic basis for these diseases. Combining Next Generation Sequencing with a new PCR-based enrichment method generates a robust and cost-effective workflow for deeper interrogation of targeted genomic regions of interest for specific applications. Here, we report the use of Next Generation Sequencing with PCR-based enrichment to extract target regions from tissue DNA. We present an optimized and flexible workflow for library construction post-PCR enrichment to enable PCR and sequencing on a Next Generation Sequencing platform. We demonstrate that this pipeline provides a useful solution for targeted resequencing applications.

INTRODUCTION

The identification of genetic variants and mutations associated with complex diseases requires the development of a robust and cost-effective approach for systematic resequencing of candidate regions in the human genome. When combined droplet-based PCR enrichment approach (Figure 2), the scalable throughput of the SOLiD™ System enables deep sequencing of target genomic regions of interest. The method employed by the Raindance Enrichment Solution can amplify regions representing up to 20 Mb of genomic sequence for parallel variant screening in a large number of genes and large numbers of samples. Post-enrichment material is then purified and incorporated into the SOLiD™ System workflow for library generation, templated-based preparation, and ligation-based sequencing (Figure 1). The inherent validity and specificity for the Raindance enrichment solution coupled with the high throughput of the SOLiD™ sequencing platform provides an integrated approach to targeted resequencing that is particularly suited for medical and cancer research as well as for follow-up Genome Wide Association Studies (GWAS).

MATERIALS AND METHODS

Resequencing Targets and Primer Design

The Raindance Oncology Panel consists of 162 genes that contain driver mutations in a number of common cancers. Target regions include exons, splice junctions, 5'- and 3'-UTRs, and promoters in these genes. 3,979 amplicons were designed to cover the targets at 100% success rate with the total amplicon sequences of 1.5Mb.

Primer Droplet Library Generation

Purified and non-purified primers were prepared at equal concentrations for each of the 3,979 amplicons. The primers were reformatted into droplets in a serial process and pooled into a single droplet library (Figure 2). Aliquots of the droplet library were prepared for use on the RDT 1000.

Targeted Enrichment and Library Preparation

Enrichment was performed according to the Raindance RDT 1000 Sequence Enrichment Assay Manual. In short, for each experimental condition, 10 ng DNA from HapMap NA18858 (Yoruba URA) was fragmented to a size of 2-4 kb using the Covaris™ S2 System (Covaris, Inc.) according to the DNA mini-TUBE™ - blue protocol from Covaris Inc (www. http://covarisinc.com/support-protocols.html). The PCR template mix containing the fragmented DNA and PCR reagents was heated to 95°C for 5 min before adding the primers and primer library. Denatured primers were added into 0.2 μl PCR tubes and amplified using 5 cycles of PCR. Amplification products were recovered by breaking the emulsions and amplification purification was performed using a MinElute PCR Purification Kit (Qiagen). Quantitative and qualitative analysis of the amplification products was performed on an Agilent 2100 Bioanalyzer (Figure 3).

Up to 750 ng of enriched DNA was purified using Invitrogen’s 2x Gel Elute™ gel according to the DNA Purification Using 2x Gel Elute™ Aparatus protocol in the 2x Gel Elute™ Technical Guide. During target amplification selection, fractions containing the amplicons were collected during 2-4 kb intervals of enrichment and concatenation on enrichment performance

 PCR-based targeted sequence enrichment for next generation sequencing platform

R. Padilla1, A.B. Shah1, L. Z. Pham2, J. Ni2, M. Matvienko1, N. Hoag1, K. Li1, J. Ziegle1

1Life Technologies, Foster City, CA, United States; 2RainDance Technologies, Inc., Lexington, MA, United States.

RESULTS

Figure 1. SOLiD™ System PCR-based Targeted Resequencing workflow

Figure 2. Microdroplet PCR-based Targeted Enrichment

Figure 3. Relative amplification abundance and length for Raindance Oncology Panel

Figure 4. Strategy for post-enrichment preparation for sequencing sensitivity and specificity

Figure 5. Percentage of Bases in the Targeted Regions Versus Depth of Coverage for All Enrichment Samples

Figure 6. Effect of amplon gel purification and concatenation on enrichment performance

CONCLUSION

The SOLiD™ System and the Raindance Enrichment Solution provide a powerful PCR-based targeted resequencing solution for detection of genetic variation. Our results demonstrate highly specific PCR-based enrichment of approximately 1.5 Mb target genomic regions, based on the amplicon abundance and length (Figure 3B) as well as enrichment specificity for all samples (Figure 6).

Our sample preparation strategy sought to enhance the existing SOLiD™ System PCR-based targeted enrichment workflow, by evaluating the effect of amplon gel purification and concatenation time of enrichment efficiency (Figure 4). Amplon gel purification results in higher specificity of the reads to the desired target, as shown by comparison of unconcatenated, gel-purified (GP, No Ligation) and non-gel-purified (OA, No Ligation) mapping profiles (Figure 5A and 5B). The findings also show no significant improvement in target specificity when increasing concatenation reaction time from 30 min to overnight (GP, 30 min ligation vs. GP. 0 min ligation), but we can see the potential that other parameters, such as increasing the units of ligation, may further enhance concatenation efficiency. Optimal enrichment efficiency was achieved by post-enrichment gel purification of the amplification products followed by end-repair and concatenation performed prior to library construction (Figure 6).

For accurate detection of genetic variants, the extent of coverage for the target regions was assessed for all of the enrichment samples. When 6 all samples were evaluated as a function of average coverage, 98% or more of the target bases were covered by at least 1 read, while 91% or more of the target bases were covered by at least 30 reads (Figure 5A). Evaluation of all 6 samples as a function of normalized coverage produced similar results (Figure 5B). Coverage profile characteristics were highly reproducible under the same sample preparation conditions (data not shown). Thus, the specificity and sensitivity of the Raindance Technology coupled with the accuracy and throughput of the SOLiD™ System is ideal for detection of genetic variants.

In order to determine sensitivity, specificity, and concordance of SNP detection, SNPs identified in this study were compared to genotypes in the HapMap database for the same sample. Classification of SNPs is outlined in Table 1. The total number of SNPs identified in targeted regions by the SOLiD™ System is shown for the 6 enrichment samples (Table 1). For all samples, approximately, 20% of the SNPs identified in the targeted regions are novel.

Reported sensitivity and specificity values for the gel purified, concatenated samples are shown (Table 2), as these samples showed the greatest enrichment specificity. For homozygous and heterozygous SNPs for each samples, sensitivity and specificity were greater than 98% and 100% respectively.

Table 1. SNP Calls for Enriched Samples Sequenced by the SOLiD™ System

Table 2. SNP Discovery Statistics in Enriched Samples

ACKNOWLEDGEMENTS

We thank Shannon Han for data analysis.