

Parallel Production of Barcoded Amplicon Libraries for 454 Sequencing using Individual and Multiplex PCR in Nanoliter Volumes

Andrew P. May¹, Jun Wang¹, Fiona Kaper¹, Megan J. Anderson¹, Peilin Chen¹, Victor Lonsberry², Casey Matthews², Joseph Boland², Min Lin¹, Martin Pieprzyk¹, Robert C. Jones¹

¹Fluidigm Corporation, 7000 Shoreline Court, Suite 100, South San Francisco, CA 94080
²SAIC, Laboratory of Molecular Technology, Gaithersburg, MD

Introduction

PCR remains the gold standard in terms of uniformity and sensitivity for enriching selected regions of interest for sequencing. However, traditional PCR is difficult to scale and requires significant quantities of DNA for each amplification reaction. We have developed a novel microfluidic platform, the Access Array™ system, to address the problems associated with scaling up PCR to the throughput required for second-generation sequencers. The Access Array system combines 48 samples systematically with 48 primer pairs to assemble 2304 PCR reactions using only 50ng input DNA per sample. In addition, we have developed a primer design strategy that incorporates sample-specific barcodes and 454 sequencing adaptors into each sample, removing the need for library preparation before sequencing.

Access Array™ System

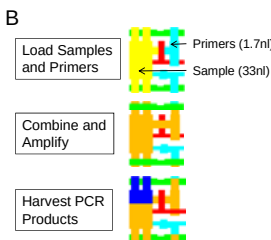
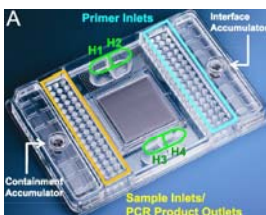
The Access Array system is centered on the 48.48 Access Array Integrated Fluidic Circuit (IFC) (Figure 1A). This is a microfluidic chip that systematically combines 48 sample inputs with 48 primer inputs to make all possible 2304 combinations of samples and primers. The chip is mounted on an SBS-compatible carrier, for simple loading of reagents with an 8-channel pipettor. Once samples and primers have been loaded, and combined into the 2304 reactions, the chip undergoes thermal cycling to amplify regions of interest from the samples. After thermal cycling has been completed, reaction products are recovered on a per-sample basis using a pump system built into the chip (Figure 1B).

The 48.48 Access Array IFC requires only 5 µl PCR master mix and 50ng input Human genomic DNA per sample. Genomic DNA can be used directly and requires no additional fragmentation or modification before loading into the chip.

The Access Array system (Figure 1C) comprises two IFC-controllers, one pre-PCR for loading samples and primers, and one post-PCR for harvesting PCR products, and a thermal cycler. When used in conjunction with the Biomark™ real-time PCR system, the progress of individual PCR reactions can be monitored during amplification.

When used in combination with primers designed for Amplicon tagging, the Access Array produces sequencer-ready amplicon libraries that can be introduced to emulsion PCR for 454 sequencing. The output from an Access Array is well matched to the capacity of the 454 FLX Sequencer with Titanium chemistry. 48 samples prepared on an Access Array can be sequenced at an average of 50x coverage on 1/8 of a picotiter plate.

Figure 1: The Access Array System



Sequencer-ready PCR products

We have developed an amplicon-tagging method (Figure 2) that enables direct attachment of sample-specific barcodes and sequencer-specific tags to PCR products generated in the Access Array. Region-specific primers are tagged at the 5' end with universal sequence tags (Figure 2A). Barcode primers comprise 454 sequence tags, a barcode sequence, and the universal tags (Figure 2B). Barcode primers are added to the samples before loading onto the Access Array IFC. The Access Array IFC combines the region-specific primers, samples and barcodes in each reaction (Figure 2D) to generate a fully tagged, barcoded PCR product (Figure 2C) in each chamber.

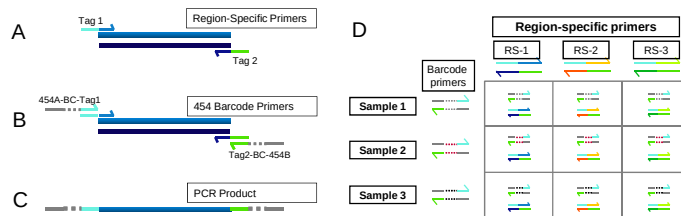
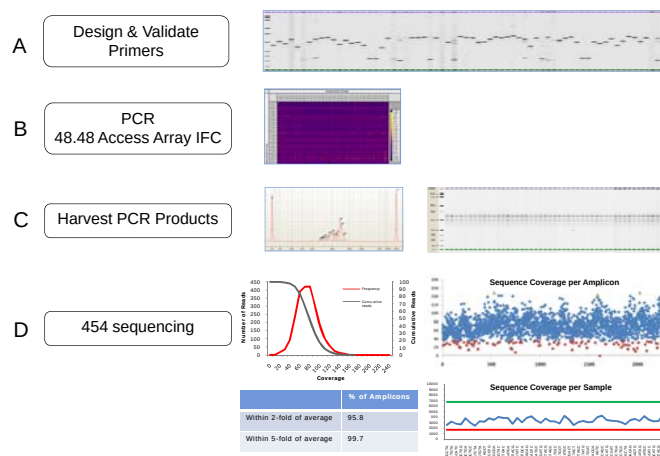


Figure 2: Amplicon-tagging to introduce sequencer-specific tags and sample-specific barcodes

High quality sequence data



48 Primer sets, designed against regions of interest in prostate cancer, were designed and validated by PCR (A). These 48 regions were amplified from 48 HapMap samples (normalized to 50ng/ul) using the amplicon tagging protocol and run on a 48.48 Access Array IFC (B). Harvested PCR products from all 48 samples (C) were pooled and run on a 454 FLX sequencer. Representation of each was even, with greater than 95% represented within 2-fold of average coverage.

Increasing enrichment per sample

With multiplexing of primer pairs in each reaction, there is potential for increasing the number of regions enriched per sample, from the same quantity of input DNA, and at the same cost per sample (around \$7). Primers were designed to amplify the exons of the EGFR and MET genes (Figure 4A). These primer sets were then used to amplify regions from a standard human genomic DNA sample as 48 individual primer sets or in 5 groups of multiplex primers. Individual PCR reactions were run using the amplicon tagging protocol. Multiplex PCR was run in a 2-step reaction. The first step involved amplification of the region-specific multiplex pools, the second step involved addition of the outer primer pairs to harvested PCR products from the first step, followed by 10 cycles of PCR in a PCR tube to assemble the tagged, barcoded samples. Samples were sequenced directly on a 454 FLX Sequencer using Titanium Amplicon chemistry with no additional purification. Amplicon representation was normalised relative to the average coverage for each sample and plotted for both protocols (Figure 4A). 90% of 10-plex reactions are represented within 2-fold of average. 93% of 1-plex reactions are within 2-fold of average. Representation of each amplicon within the individual and multiplex PCR reactions was very similar.

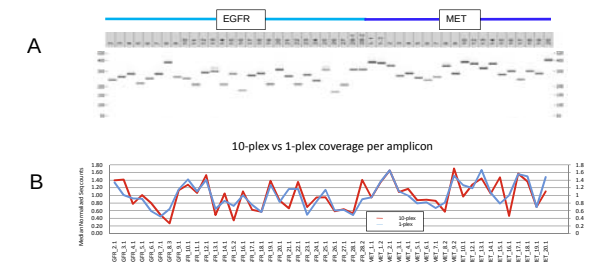


Figure 4: Comparison of individual and multiplex PCR reactions run on the Access Array system.

Conclusion

- The Access Array System provides a simple, easy-to-use method for PCR-based target enrichment from multiple samples in parallel using only 50ng human genomic DNA per sample for 48 PCR reactions
- We have developed an amplicon-tagging protocol that enables direct labeling of PCR products with sample-specific barcodes and sequencer-specific tags
- Data from sequencing experiments demonstrates that amplicon representation within and between samples is highly uniform
- We have demonstrated the potential for multiplexing primer pairs within PCR reactions on the Access Array system. Representation of each amplicon in the multiplex reaction is similar to that within individual PCR reactions.



Fluidigm Corporation
 7000 Shoreline Court, Suite 100 • South San Francisco, CA 94080
 Toll-free: 1.866.FLUIDLINE
 www.fluidigm.com

We would like to acknowledge 454 sequencing services for assistance with sequencing