Parallel Preparation of Targeted Resequencing Libraries from 480 Genomic Regions Using Multiplex PCR on the Access Array™ System

Fiona Kaper, Jun Wang, Megan J. Anderson, Peilin Chen, Min Lin, Martin Pieprzyk, Robert C. Jones, Andrew P. May

Fluidigm Corporation, 7000 Shoreline Court, Suite 100, South San Francisco, CA 94080

Introduction

Next generation sequencing platforms have dramatically reduced sequencing costs. However, it currently remains too expensive to routinely resequence entire human genomes in order to discover genetic variants or somatic mutations underlying tumorigenesis. Therefore, a need exists for multiplexed, targeted amplification methods that allow for the analysis of multiple genomic regions in large cohorts. Available targeted enrichment technologies are either aimed at the capture of regions of interest from a single sample, exhibit uneven representation or require significant amounts of input material. The novel microfluidic platform, the Access Array™ system, combines 48 samples with 48 primer sets resulting in 2,304 simultaneously occurring PCR amplifications requiring as little as 50ng DNA per sample. PCR products generated on the Access Array system can be used for sequencing on all next-generation sequencing platforms, including 454 GS FLX and Illumina GAII. To increase coverage and throughput, PCR reactions can be multiplexed within Access Array IFCs generating up to 480 amplicons per sample.

Access Array™ System

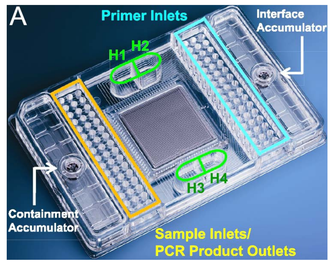
The Access Array system is centered on the 48.48 Access Array Integrated Fluidic Circuit (IFC) (Figure 1A). This is a microfluidic chip that systematically combines 48 sample inputs with 48 primer inputs to make all possible 2304 combinations of samples and primers. The chip is mounted on an SBS-compatible carrier, for simple loading of reagents with an 8-channel pipettor. Once samples and primers have been loaded, and combined into the discrete 2304 reactions, the IFC 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).

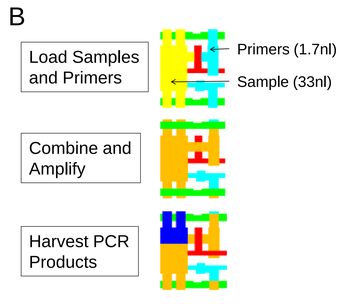
The 48.48 Access Array IFC requires only 5 ul PCR master mix and 50ng input Human genomic DNA per sample. Genomic DNA can be used directly and requires no additional fragmentation or modification before loading into the chip

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 IFC is well matched to the capacity of the 454 FLX Sequencer with Titanium chemistry. 48 samples prepared on an Access Array IFC can be sequenced at an average of 50x coverage on 1/8 of a picotiter plate.

Figure 1: The Access Array System





C

Sequencer-ready PCR products

We have developed a multiplexed 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 IFC. Region-specific primers are tagged at the 5' end with universal sequence tags (Figure 2A). 48 pools of up to 10 Region-specific primer pairs are pooled and loaded into the assay inlets of the Access Array IFC. The 48.48 Access Array IFC then combines the region-specific primer pools with the samples in each reaction. After harvesting the PCR product pools from the 48.48 Access Array IFC, a second PCR is carried out on a 96-well plate with barcode primers comprising 454 sequence tags, a barcode sequence, and the universal tags (Figure 2B). The final result is a fully tagged, barcoded PCR product (Figure 2C).

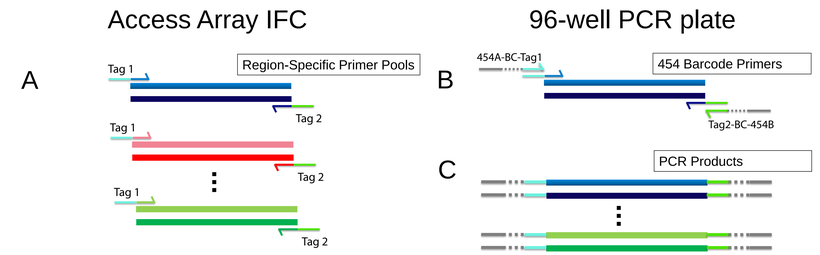
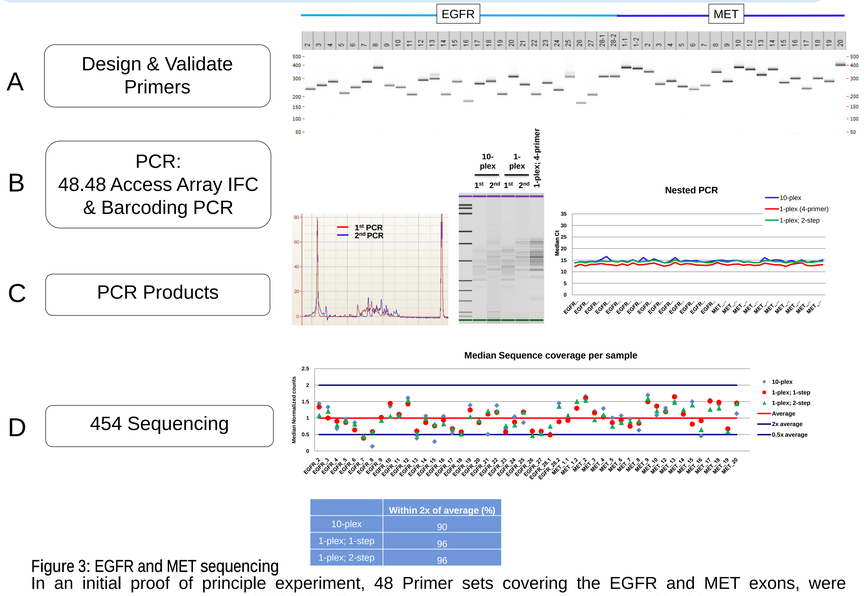


Figure 2: Amplicon tagging to introduce sequencer-specific tags and sample-specific barcodes

High quality sequence data



	Within 2x of average (%)
10-plex	90
1-plex; 1-step	96
1-plex; 2-step	96

Figure 3: EGFR and MET sequencing

In an initial proof of principle experiment, 48 Primer sets covering the EGFR and MET exons, were designed and validated by PCR (A). The 48 regions were amplified from a HapMap control sample (normalized to 50ng/ul) using 5 pools of 9 to10 primer pairs each, following the strategy outlined above (Figure 2). The remaining 43 assay inlets on the Access Array IFC were filled with buffer. A comparison was made to singleplex, 2-step PCR and our singleplex, 1-step amplicon tagging strategy. Final PCR products (C) were analyzed on an Agilent BioAnalyzer 2100 and amplicon representation was assessed with a nested Q-PCR. The PCR products were subsequently run on a 454 FLX sequencer. Representation of each PCR product was even, with 90% represented within 2-fold of average coverage (D).

480 Cancer Gene Exons

To maximally utilize the capacity of an 48.48 Access Array IFC, we carried out multiplexed amplifications of a set of commonly mutated cancer gene exons across 48 genomic DNA samples. In initial experiments, 580 tagged primer pairs were validated in individual PCR reactions in 384-well plates. 480 primer pairs that produced a single band of the correct size, as determined on a Caliper LabChip system, were selected for multiplex PCR experiments. Primer pairs were grouped in sets of 10 based on their expected PCR product sizes, resulting in 48 primer pools of 10 primer pairs each. Multiplex PCR was carried out following the strategy outlined in Figure 2. The resulting products are 48 uniquely barcoded PCR product pools, each comprising 480 amplicons derived from one sample, that are ready for sequencing. Harvested PCR product pools were analyzed on an Agilent BioAnalyzer 2100 (Figure 4A) and PCR product representation was assessed with a nested Q-PCR for a subset of amplicons (Figure 4B). The PCR products were subsequently run on a 454 FLX sequencer. Representation of each PCR product was even, with 90% represented within 5-fold of average coverage.

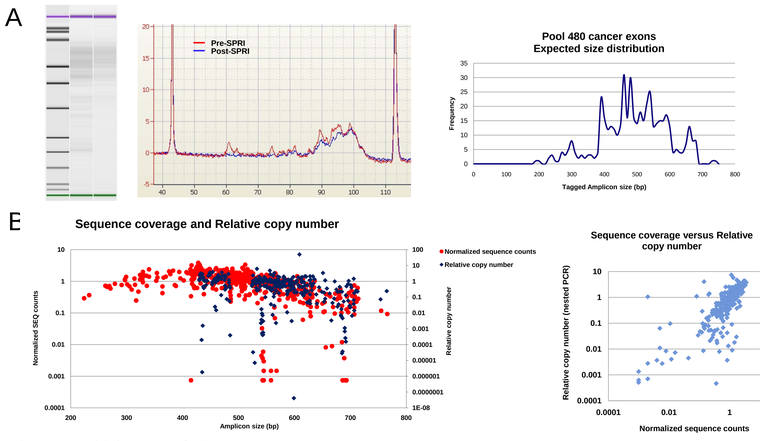


Figure 4: Multiplex PCR of 480 cancer gene exons.

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 up to 480 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 PCR product 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

Acknowledgement

We would like to acknowledge 454 Life Sciences Sequencing Services for assistance with sequencing.