Variant detection with an ultra-multiplexed 20,000-amplicon CleanPlex® NGS panel

PARAGON GENOMICS

Yang Lily Liu, Kathryn Pendleton, Lifeng Lin, Lucie Lee, Jeffery Liu, Guoying Liu, Zhitong Liu

Paragon Genomics, Inc, 3521 Investment Blvd Suite 1, Hayward, CA 94545, USA

Abstract

Next-generation sequencing (NGS) has become a key tool for screening large numbers of genes and cancer mutations because of its ability to simultaneously sequence a large number of targets. To most effectively use the power of NGS, multiplexed targeted sequencing promotes reads only in the regions of interest and can provide highly comprehensive diagnostic methods for tumor profiling and patient treatment identification. Unique challenges for these methods call for a highly efficient, accurate and robust method for the unbiased enrichment of tens of thousands of target regions. However, an increase in multiplexity can also increase the risk of non-specific primer-primer interactions, decrease uniformity, and increase GC bias. Here we present a highly-multiplexed technical panel containing ~20,000 regions, covering 364 genes. Using the CleanPlex® PCR-based method, this unoptimized amplicon panel exhibits ~93% uniformity at 0.2X mean bias, limited GC bias, and mutation calling at low allele frequency. Paragon Genomics CleanPlex® technology demonstrates that it is capable of creating high quality amplicon libraries with high uniformity, low GC bias, and sensitive variant *calling even in ultra-multiplexed libraries with ~20,000 amplicons with a workflow under 3-hours.*

Workflow

The Paragon Genomics CleanPlex® workflow is designed for straightforward and rapid preparation of NGS libraries for the Illumina platform. Compared to other methods that can take up to 4 days, the CleanPlex® workflow only requires 2.5 hours with 30 minutes of hands on time.

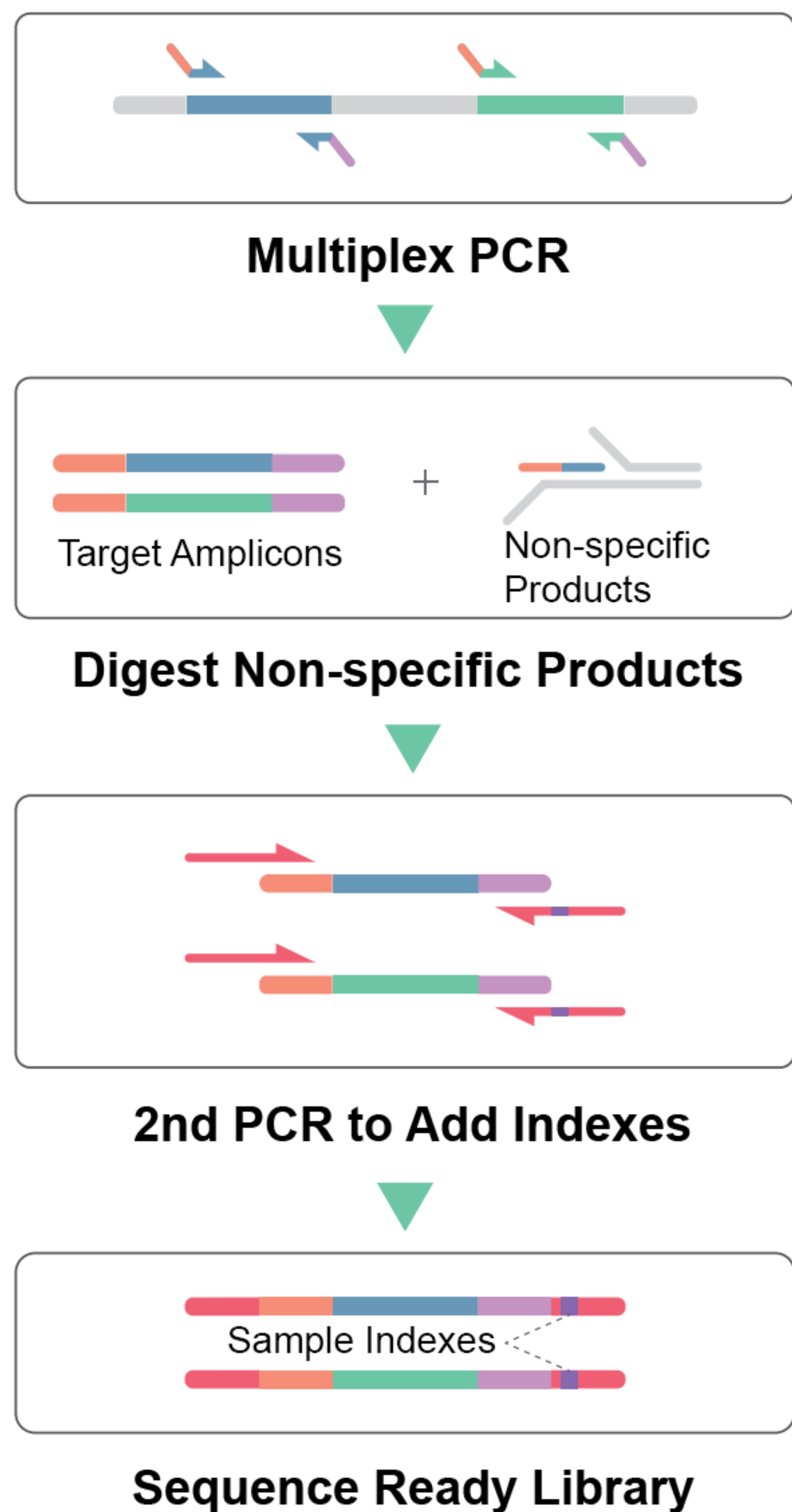

Figure 1: The Paragon Genomics Library Preparation workflow begins with amplification of target sequences through multiplex PCR using oligonucleotides that contain a partial adaptor sequence. Primer-dimers, heteroduplexes, and other side products are then degraded and removed from the reaction solution using the CleanPlex® Digestion Reagent. Finally, Illumina adaptor sequences and sample indexes are added to the library through a second round of PCR, which anneals to the partial adaptor sequence added to each amplicon during the multiplex PCR step.

Panel Design

We designed a panel targeting the coding exon regions of 364 genes, with 15 bp paddings at exon-intron junctions. The amplicons are designed using our internally developed primer design pipeline, Paragon Designer™, which selects amplicons based on primer thermodynamic properties, primer binding specificity and inclusivity of primers, and uniqueness of the amplicon insert sequences. Roughly 20,000 amplicons were designed to cover 985,512 bp/ 1,082,233 bp (91.06%) of the targeted regions. The primers are divided into two pools according to their compatibility. The cumulative length of amplified fragments is 2,237,047 bp. The amplicon length ranges from 105 bp to 120 bp, with an average of 113 bp.

Gene Coverage, Mapping, On-Target Rate, and Uniformity

Performance Statistics	5% mix	10% mix
% Mapped Reads	71.0%	79.6%
% On-Target	96.8%	97.1%
Uniformity (0.2X mean)	93.8%	92.9%

Table 1: The statistics for the percentage of mapped reads, percentage of reads on target, and the average uniformity (the percentage of amplicons above 0.2X the mean read depth) for the ~20,000 amplicon panel prepared by Paragon Genomics' CleanPlex® Technology.

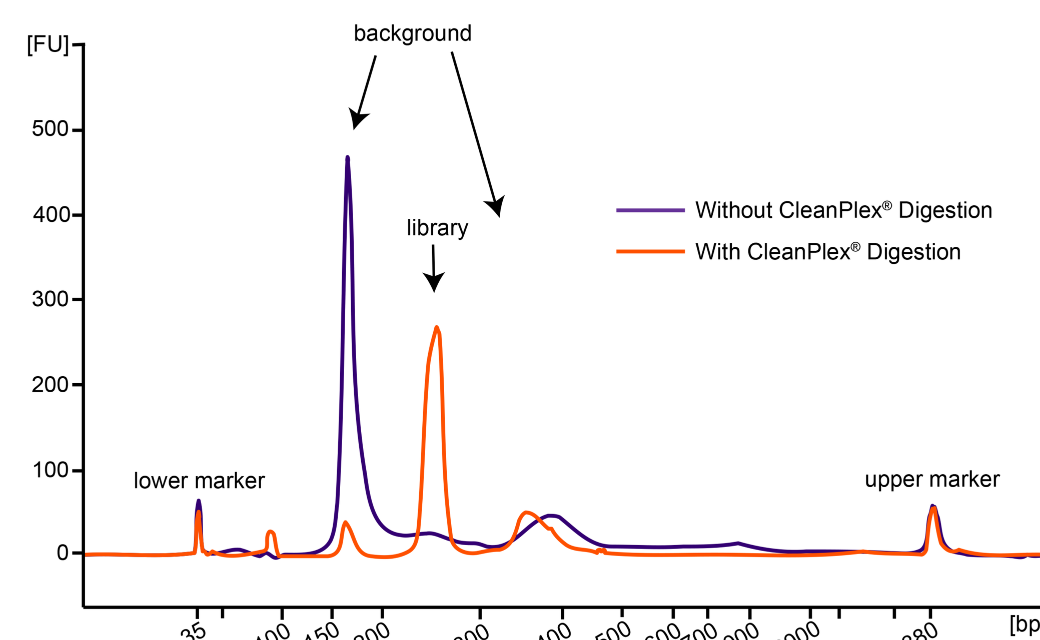

Figure 2: Comparison of Agilent BioAnalyzer traces of highly multiplex panel (~20,000 amplicons) treated (orange plot) and not treated (purple plot) with CleanPlex® Digestion Reagent. Two pools of multiplexed primers are used for the initial PCR with 10ng of genomic DNA per pool. The products are combined after the first bead clean-up. The proprietary CleanPlex® Digestion Reagent is essential for obtaining the desired library and reducing undesired side products formed during the mPCR amplification of target sequences.

GC Bias

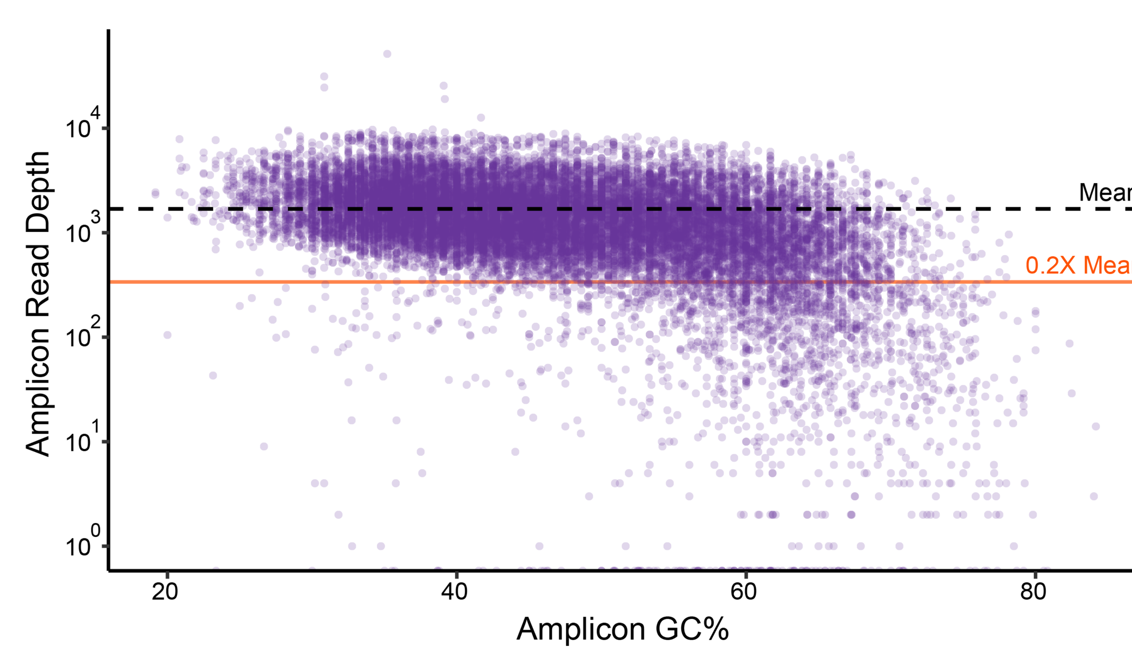

Figure 3: Amplicon Read Depth vs. Amplicon GC% Content of a library of the ~20,000 amplicon panel. The dashed line represents the mean amplicon read depth. The orange line represents the 0.2X mean value. For amplicons with >55% GC content, 79% are above 0.2X the mean reads. Future optimization, such as primer re-design, could enhance any under-performing amplicons.

Variant Detection

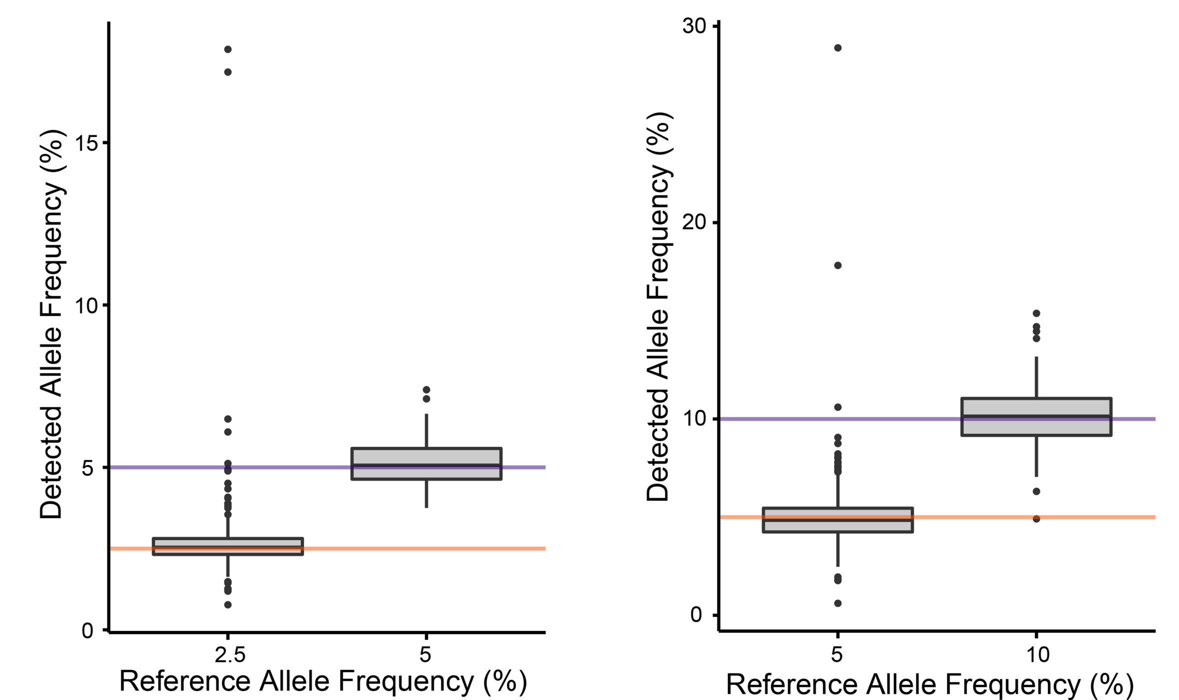

Figure 4: 10ng of genomic DNA (gDNA) mix of 5% NA18507 in NA12878 (left) or 10% NA18507 in NA12878 (right). The purple line represents the expected allele frequency (AF) of the homozygous mutations for the samples. The orange line represents the expected AF of the heterozygous mutations. The boxplot represents the AFs called for the known mutations based upon the CleanPlex®-prepared libraries.

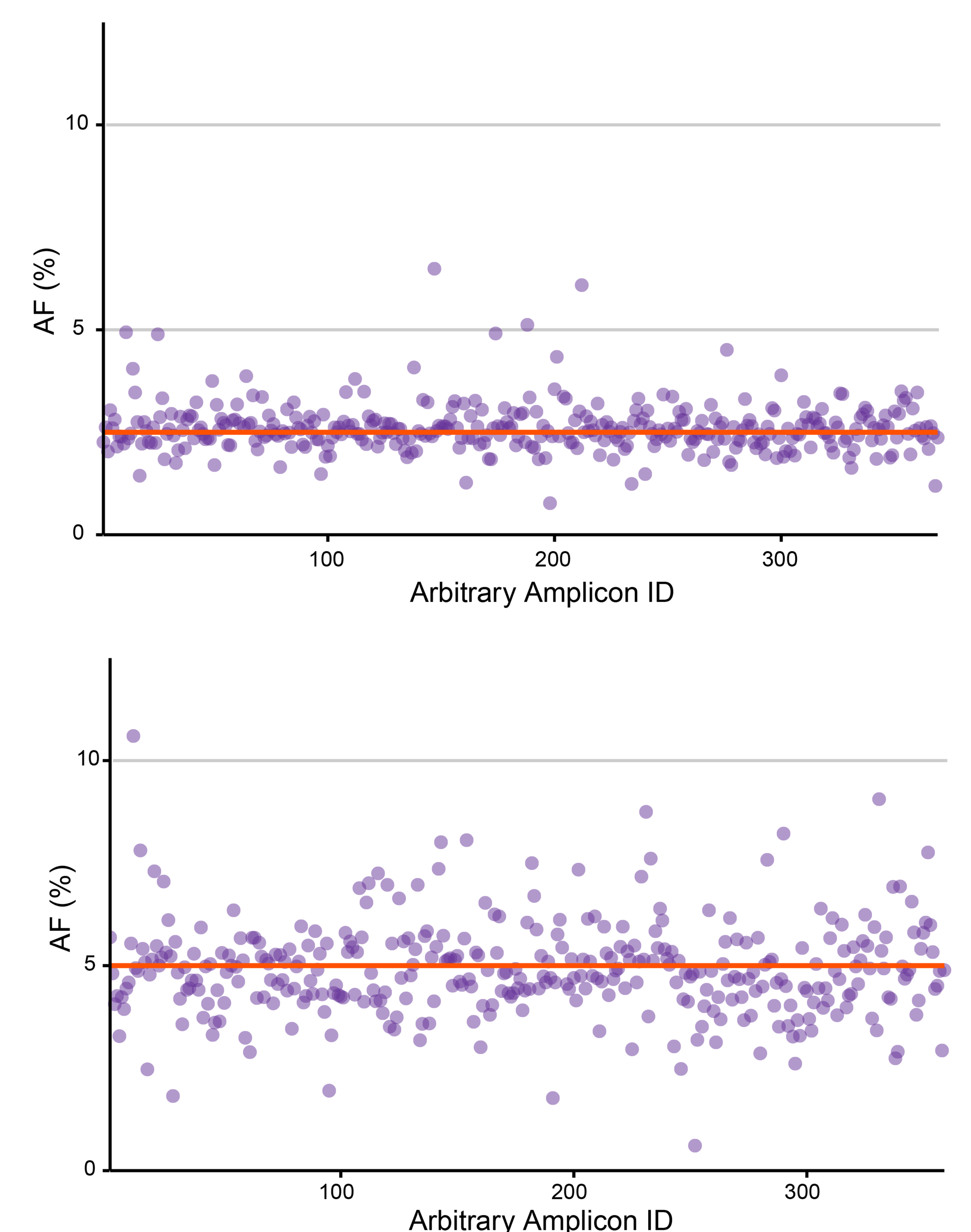

Figure 5: 10ng of genomic DNA (gDNA) mix of 5% NA18507 in NA12878 (top) or 10% NA18507 in NA12878 (bottom). 383 heterozygous mutations are expected, for the respective frequencies of 2.5% and 5%. The dots represent the detected allele AF of the ~20,000 amplicon panel. The orange line depicts the expected frequency. The 5% gDNA mix had an average read depth of ~9000 per amplicon. The 10% gDNA mix had an average read depth of ~1700 per amplicon. For each plot, four additional points are above the 12.5% AF (not shown).

Detection Rate	mAF	Total Mutations	% Detection
5% mix - Heterozygous	2.5%	383	96%
5% mix - Homozygous	5%	103	97%
10% mix - Heterozygous	5%	383	94%
10% mix - Homozygous	10%	103	96%

Table 2: For the minor alleles in 5% NA18507 in NA12878 and 10% NA18507 and NA12878, the detection frequency calculated for both the heterozygous (2.5% and 5% minor allele frequency) and homozygous (5% and 10% minor allele frequency) variants. Future optimization, such as primer re-design, could enhance any under-performing amplicons.

Conclusions

Paragon Genomics made a technical panel that demonstrates the CleanPlex® technology is capable of multiplexing up to 20,000 amplicons with no special equipment besides a thermocycler. Without any amplicon optimization, this ~20,000 amplicon panel demonstrates:

- Proprietary multiplex PCR reagents and expert primer design result in >90% uniformity and limited GC bias.
- Reliable detection of allele frequency of 2.5%.
- A highly multiplexed CleanPlex® panel produces high mapping and on-target rates with low input DNA.
- CleanPlex® Digestion Reagent significantly reduces the background products, such as adaptor-dimers, prior to sequencing.

This panel sets up the core targets of a Tumor Mutational Burden Panel that we're currently working on. The TMB panel is expected to cover ~450 genes. Final optimization will be performed for the final larger panel.

FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

© 2018 Paragon Genomics, Inc. All Rights Reserved. All trademarks are the property of Paragon Genomics, Inc. or their respective owners.

discover MORE with less™