Application Note

Multiplex PCR Targeted Enrichment Solution on the G4 Sequencing Platform

Abstract

This application note demonstrates the compatibility of Paragon Genomics CleanPlex® library kits with the G4™ Sequencing Platform by Singular Genomics. Here, we showcase the compatibility of the G4 across two Ready-to-Use panels of various amplicon panel sizes, pool quantity, and application focus, and demonstrate robust performance across both panels, which highlights seamless integration of the G4 Platform with CleanPlex library preparations.

Introduction

To overcome shortcomings in existing techniques and to bridge the gap between amplicon and hybrid capture-based methods, CleanPlex® technology was developed to provide a better tool to prepare NGS libraries for targeted sequencing. This technology is an ultra-high multiplex PCR-based target enrichment method that features a proprietary background cleaning chemistry, which allows tens of thousands of amplicons to be pooled in one reaction to target genomic regions spanning up to megabases in total size. CleanPlex technology has a simple workflow, with a low input requirement and yet produces high on-target and uniform NGS libraries that allow for efficient use of sequencing reads with accurate and sensitive detection of variants. Coupled with a highly advanced proprietary multiplex PCR primer design algorithm (ParagonDesigner™), CleanPlex chemistry provides fast and high quality custom targeted panels for a variety of applications such as oncology, inherited diseases, infectious disease, agrigenomics, and more.

The versatility of CleanPlex chemistry, which only requires a simple swap of sequencer-specific adapter primers, allows it to stay current with the ever-dynamic sequencing landscape. In this technical note, we present data to demonstrate the compatibility of CleanPlex technology with the Singular Genomics G4 Sequencing Platform.

The G4 Platform is a highly versatile benchtop sequencer that leverages a 4-color rapid sequencing by synthesis (SBS) chemistry with advanced optics and fluidics engineering to provide single-day turnaround times across all applications. By combining fast run times and the ability to run up to 4 flow cells, each with 16 independently addressable lanes, the G4 enables highly efficient laboratory operations. The G4 Platform outputs FASTQ format files that integrate seamlessly with existing bioinformatics tools. Users may elect to automatically demultiplex samples on-instrument via sample indices provided by a sample sheet or off-instrument using the Singular Genomics rapid demultiplexing tool. More information about G4 specifications, such as run time, accuracy, and quality metrics, can be found on the Singular Genomics website.

CleanPlex library solutions combined with sequencing on the G4 Platform provide robust and high-quality targeted sequencing data, with 3 hours for library preparation and a single-day turnaround for sequencing. This combination allows for highly efficient and high-capacity data generation, providing a valuable option for investigators of various markets, for example, in time sensitive oncology applications.

Application Focus

To demonstrate the compatibility of Paragon Genomics' CleanPlex library preparation technology with the Singular Genomics G4 Platform, two kits with different breadth within the field of oncology were selected. Both panels fall under Paragon's previously validated 'Ready-to-Use' portfolio, but are comprised of unique attributes and workflows as part of the CleanPlex technology.

Panel	OncoZoom	Hereditary Cancer
Amplicons	601	1,447
Amplicon size	125-175bp (146 avg)	105-299bp (214 avg)
Pools	1	2
Genes Covered	65	37
Variant Types	SNV, Indels	SNVs, Indels, CNV

Table 1. Specifications for CleanPlex OncoZoom and Hereditary Cancer Panels.

The CleanPlex® OncoZoom® Cancer Hotspot Panel is a multiplex PCR-based targeted resequencing assay designed for rapid detection of somatic mutations across over 2,900 hotspot regions of 65 oncogenes and tumor suppressor genes with known cancer associations. 601 amplicons have been designed to cover the hotspot regions including SNVs and indels. Altogether, this panel is optimized to deliver data with high on-target rates and high coverage uniformity to ensure efficient use of sequencing reads and provides a robust measure of general compatibility for the G4 Sequencing Platform.

The CleanPlex® Hereditary Cancer Panel v2 is a targeted resequencing assay designed for analyzing genes associated with an increased risk of developing hereditary cancers. The panel was curated to target the entire coding region of 37 genes which covers single nucleotide variants (SNVs) and insertion/deletion mutations (indels) or copy number variations (CNV) of interest. Furthermore, it also detects hotspot mutations rs12516 and rs8176318 in the BRCA1 3' UTR and structural rearrangement of exons 1-7 in MSH2 (Boland inversion). As a complement to the OncoZoom hotspot panel, this panel was chosen for its larger amplicon panel size consisting of two-pools, a tiled design, and full gene coverage, to assess the detection potential of a more challenging set of targets.

SINGULAR GENOMICS

PARAGON GENOMICS

Materials and Methods

Library Preparation Workflow

For both panels, 8 replicates were processed using 15ng standard human genomic DNA per pool as described in their respective User Guides. Briefly, CleanPlex technology has a streamlined single-tube workflow that is performed in 3 steps and in 3 hours **(Figure 1).**

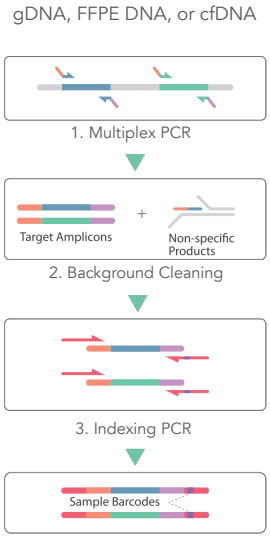

Figure 1: CleanPlex Technology workflow featuring a 3-step, 3-hr workflow.

(1) Multiplex PCR (mPCR): Targets of interest are amplified in a multiplex PCR reaction with primer pairs designed using our proprietary ParagonDesigner™ algorithm. The mPCR mix is specifically formulated to minimize GC bias and improve uniformity.

(2) Background cleaning: Primer-dimers, non-specific PCR products, and complex molecular-debris are biochemically removed in a digestion reaction.

(3) Indexing PCR: Each library is barcoded with platform specific sample indices and amplified in a PCR reaction. The resulting libraries are ready for sequencing. Singular Genomics Unique Dual Index Primers (Cat #700,134) were used directly in the 2nd PCR reaction and treated exactly as described for Illumina Index Primer workflow in the User Guide.

Sequencing and Analysis

Libraries were visualized and quantified using an Agilent Bioanalyzer 2100 with a High Sensitivity DNA Kit (Agilent, Cat #5067-4626). Final library quantification was determined via Qubit (ThermoFisher, Cat #Q33238). Libraries were pooled and sequenced on the Singular Genomics G4 Platform using an F2 flow cell with sequencing format of 2 x 150 bp (plus 12 bp dual indices). 191M read pairs were generated with 91.8% and 92.8% of base calls ≥ Q30 for Read 1 and Read 2, respectively. Raw reads were analyzed using standard analysis workflow and internal pipeline details described in both User Guides. The raw reads were trimmed using Cutadapt and then downsampled to mean coverage of between 800X-5000X for OncoZoom and 300X-2000X for Hereditary Cancer using Seqtk. This data was then processed using internal pipelines generated as outlined in both User Guides. For each panel, alignment was performed with BWA to reference genome GRCh37 (hg19), and variant calling was performed via VarDict variant caller.

Results

Library quality was first assessed by examining the Bioanalyzer traces of the libraries generated following the respective workflows. Two Illumina-ready libraries were also processed alongside the Singular Genomics material for direct benchmark assessment of the library trace fragments.

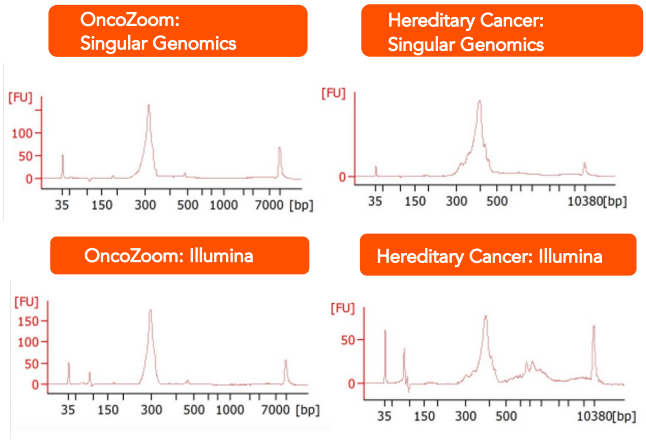

Figure 2: Bioanalyzer trace of libraries generated with Singular Genomics UDIs and Illumina dual indices processed in parallel. Traces show expected peak location and distribution.

Figure 2 shows representative examples of fragment analysis traces for both panels generated for sequencing on the G4 Platform, and comparison to the corresponding Illumina library counterpart. In both cases, libraries are of the expected shape and position, with a slight shift to account for the additional length of Singular Genomics unique dual indices (+24bp total, 12bp each side) compared to the Illumina combinatorial dual indices used (+16bp total, 8bp each side).

After confirming the library peak distribution, all libraries generated for Singular Genomics were processed and sequenced on the G4 Platform (Illumina library counterparts were not sequenced given previous validations). **Figure 3** shows representative examples of the read depth uniformity of these panels by plotting the amplicon read depth (in logarithmic scale) against the amplicon GC% in order to assess GC content bias, a common limitation in the NGS field that can especially impact methods involving multiplexed PCR. For both panels, no unexpected target drop out or striking bias by amplicon GC content is observed, suggesting the robust adaptations built into CleanPlex mPCR chemistry hold true with sequencing on the G4 Platform.

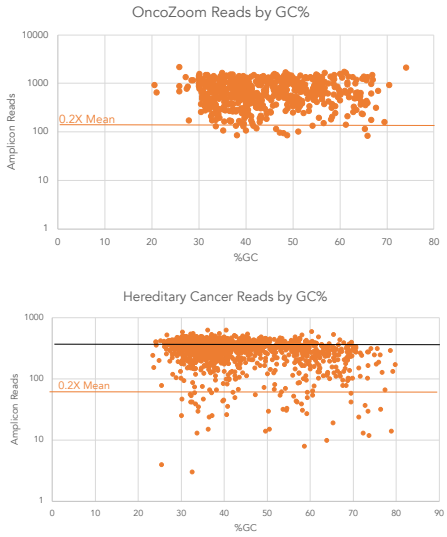

Figure 3: Performance of OncoZoom and Hereditary Cancer panels with the G4 Sequencing Platform. Amplicon distribution shows no unexpected target drop outs or obvious bias to amplicon GC content.

SINGULAR GENOMICS

PARAGON GENOMICS

Next, the uniformity, on-target rate, and primer dimer rate were quantitatively assessed for each panel. For OncoZoom, as 2000X-5000X coverage is recommended for detecting somatic mutations, downsampling was performed incrementally from 3M to 500K total reads per sample to target a range of ~800X-5000X coverage per amplicon. Similarly, for Hereditary Cancer, where 500X-1000X coverage is recommended for germline mutation detection, downsampling was performed from 1.5M to 500k total reads per sample to target a range of ~300X-2000X coverage per amplicon. In both cases, the objective was to showcase a range of coverage spanning the suggested guidelines with a slightly more robust lower end. These results are illustrated in **Table 2** and **Figure 4**.

For OncoZoom, the average uniformity, on-target rate, and primer dimer rate were 96%, 93%, and <0.02%, all of which were minimally impacted by the downsampling from 3M to 500k reads. This demonstrates a robust coverage of the panel with as low as 800 reads per amplicon and virtually no change when sequencing deeper up to 5000 reads. For Hereditary Cancer, the average uniformity, on-target rate, and primer dimer rate were 96%, 91%, and 0.41% respectively. As with OncoZoom, the downsampling again had no impact on the metrics from 1.5M to 500k reads, demonstrating 300 reads per amplicon and virtually no change when sequencing deeper up to 2000 reads.

	OncoZoom	Hereditary Cancer
Uniformity-20%	96%	96%
On-Target Rate	93%	91%
Primer Dimer	0.02%	0.41%

Table 2: Sequencing Performance metrics of panels on the G4 Platform, downsampled to 500K reads per sample. Both demonstrate high performance.

Lastly, we briefly investigated the concordance of variant calling for the two panels using a few replicate samples (**Figure 5**). For OncoZoom, to assess ability of samples sequenced on the G4 Platform to call variants of low allele frequency, samples with 1% of NA12878 spiked into 99% NA18507 were chosen to simulate a sample with a low allele frequency variant. **Figure 5 (left)** shows the plotted variant calls which demonstrates detection by the two replicates for the 1% allele frequency in the lower left via the 1% spike in of NA12878 as well as the ~50% in the middle (one variant in the two strands) and ~99% on the upper right (both variants in the two strands) detected in the majority NA18507 sample. Concordance of the two samples was high with an R^2 of 0.98. For Hereditary Cancer, concordance was evaluated with two random samples to investigate reproducibility among a much larger panel. **Figure 5 (right)** shows the variant calls for two representative samples and concordance was high with an R^2 of 0.99 demonstrating good reproducibility of variant calling.

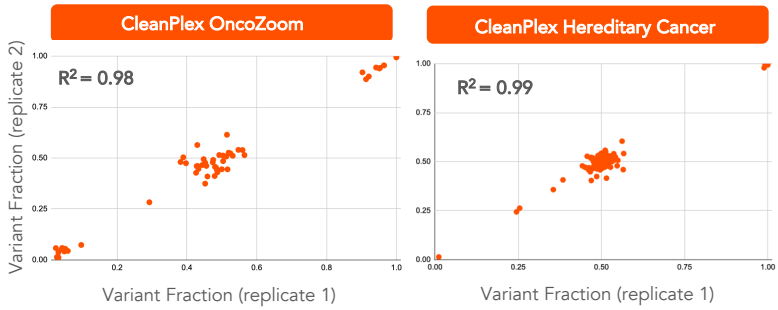

Figure 5: Variant calling concordance. Using randomly selected sample replicates, concordance was high for both panels, with R2 of 0.98 and 0.99, respectively, demonstrating strong reproducibility for variant calling. DNA spike in samples processed for OncoZoom further demonstrated low allele frequency variant detection potential using the G4 Sequencing Platform.

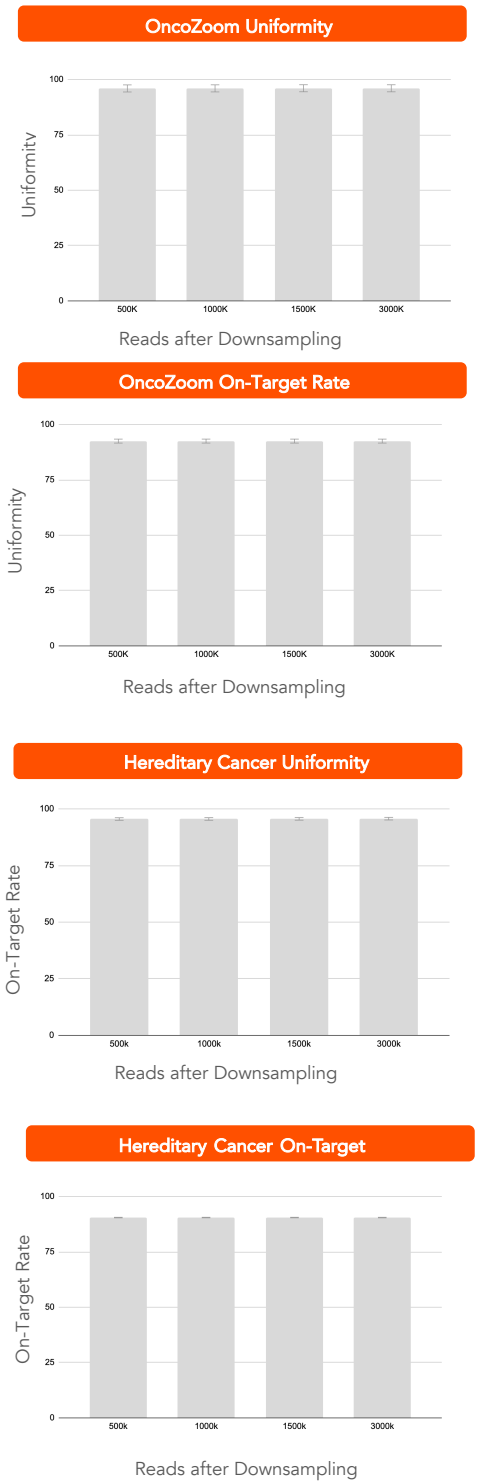

Figure 4: Quantitative performance of OncoZoom and Hereditary Cancer panels with the G4 Sequencing Platform following downsampling. Both OncoZoom and Hereditary Cancer panels had high uniformity and on target rates and showed robust performance.

SINGULAR GENOMICS

PARAGON GENOMICS

Discussion and Conclusion

In this application note, we have demonstrated the feasibility of sequencing Paragon Genomics CleanPlex targeted libraries on the Singular Genomics G4 Platform, using the Ready-to-Use OncoZoom and Hereditary Cancer panels. OncoZoom, a robust panel of cancer gene hotspot detection which serves as a gold standard for CleanPlex technology, demonstrated high performance for uniformity (96%), on-target rate (93%) and low primer dimer rate (0.02%) akin to previous validations with Illumina sequencing. Variant calling analysis was also of high concordance (R^2 of 0.98) demonstrating strong reproducibility, and use of samples with 1% DNA spike-in demonstrated the lower allele frequency variant calling potential. The larger 2-pool gene panel, Hereditary Cancer, also showed robust results akin to our previous validations in terms of high uniformity (96%), on-target rate (91%), and low primer dimer rate (0.41%) with high variant calling concordance (R^2 of 0.99). Altogether, both kits demonstrated robust performance when sequencing on the G4 Platform with a wide range of coverage.

Collectively, these experiments have demonstrated CleanPlex chemistry's compatibility with and performance on the Singular Genomics G4 Sequencing Platform and highlight the platform-agnostic nature of our workflow. The breadth of the kits chosen in this application note, their variations in the CleanPlex workflow, and the CleanPlex Technology's simple and uniform approach to various applications provide strong evidence for the technology's compatibility as a whole. We have confidence in utilizing the G4 for 'Ready-to-use' panels as well as custom panel offerings to support various research needs. With the confirmation of the G4 Sequencing Platform compatibility via a simple index primer switch, this reiterates its plug-and-play capability and operational efficiency. Paragon Genomics is proud to offer 'Ready-to-use' and custom targeted library preparation solutions that are easy to use and simple to implement for each labs' unique sequencing preferences.

Begin Your Journey with G4

Contact our sales team to to learn more about the G4 Sequencing Platform.

Website: www.singulargenomics.com
Email: care@singulargenomics.com
Call: +1 442-SG-CARES (442-742-2737)
Address: 3010 Science Park Rd, San Diego, CA 92121

Discover more with less™

Paragon Genomics, Inc. | 5020 Brandin Ct, Floor 2, Fremont, CA 94538-3140 | +1.650.822.7545
www.paragongenomics.com | support@paragongenomics.com

SINGULAR GENOMICS PARAGON GENOMICS