SINGULAR GENOMICS

APPLICATION NOTE

Targeted Sequencing Using Agilent SureSelect and the G4™ Sequencing Platform

- Rapid, accurate, and cost-efficient analysis of key variant types for oncology translational and clinical research with Agilent SureSelect products
- Seamless integration of the G4 into the Agilent SureSelect Exome V8 library preparation workflow and bioinformatics pipeline
- Highly accurate targeted sequencing data generated on the G4 Platform with a lower duplicate rate and coverage bias than the Illumina HiSeq 4000

Introduction

Targeted next-generation sequencing (NGS) has become a central component of solid tumor translational research and clinical testing by enabling detection of actionable variants with lower cost and higher sensitivity than whole genome sequencing. There is a demand for targeted NGS library preparation solutions that enable detection of SNVs (single nucleotide variants), indels (short insertions and deletions), CNVs (copy number variants), and translocations from a single customizable assay. Furthermore, there is a demand for delivering accurate targeted sequencing results with more cost-efficiency and faster turnaround time.

In this application note, we highlight the use of the Agilent SureSelect Exome V8 and a custom cancer panel with the Singular Genomics G4™ Sequencing Platform to deliver rapid, accurate, and cost-effective targeted variant detection results.

Agilent SureSelect Exome v8 and XT HS2 Kits

The SureSelect Human All Exon V8 provides comprehensive and most up-to-date coverage of protein coding regions from RefSeq, CCDS, and GENCODE. It also covers the TERT promoter and hard-to-capture exons that are omitted by other exome assays on the market. Powered by machine learning-based probe design and a new production process, SureSelect Human All Exon V8 spans a 35.1 Mb target region of the human genome with an efficient end-to-end design size of only 41.6 Mb. The SureSelect Human All Exon V8 is compatible with the streamlined SureSelect XT HS2 library preparation and target enrichment systems, which feature a fast, 90-minute hybridization protocol. The workflow is optimized for FFPE samples which are essential for cancer research. Up to 384 unique dual sample indexing enables high-throughput labs to process and sequence hundreds of samples simultaneously. The ability to generate consensus calls using MBC information from both strands (duplex MBC) significantly improves the accuracy of low VAF detection which is critical in liquid biopsy applications.

G4 Specifications for Targeted Sequencing

The G4™ Sequencing Platform is a highly versatile benchtop sequencing platform that is well-suited for demanding targeted sequencing applications. The G4 leverages a novel, 4-color Rapid sequencing by synthesis (SBS) chemistry to deliver highly accurate reads under 24 hours for typical targeted sequencing experiments (2x150bp reads). To maximize flexibility, the G4 enables users to load up to four flow cells at a time, with each flow cell comprising four fluidically independent lanes, thereby enabling sample multiplexing with or without the need for index reads. The G4 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 the 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.

Methods

Library Preparation and Sequencing

200ng enzymatically-sheared gDNA from cell line NA12878 was used as input for library preparation using Exome v8 and a custom cancer panel (targeting 519 genes) . Briefly, four libraries were prepared from gDNA using the SureSelect XT HS2 Library Prep kit and enriched with the respective panel reagents **(Figure 1)**. An aliquot of each library was processed using the Singular Genomics Compatibility Kit to introduce Singular Genomics flow cell binding sequences S1 and S2 for sequencing on the G4 (refer to Adapting Libraries for the G4™ Retaining Indices for more information). Following that step, sequencing of the Illumina libraries was performed via the HiSeq 4000, and indexed sequencing of the Singular Genomics libraries was performed using the G4 Platform with the F2 flow cell.

Analysis

FASTQ files produced by the G4 Platform and the HiSeq 4000 were aligned to hg38 via bwa mem, down sampled to 2.16M reads (custom cancer panel) or 40M reads (XTHS Exome v8) per library via Picard Downsample SAM, then % on target and 30x coverage metrics were calculated via METRICS PROGRAM, an internal tool developed by Agilent. Duplicate rates, fold-80 base penalty and AT/GC dropout values were calculated via Picard CollectHSMetrics. The Germline variants were identified via Platypus with default parameters. Variant detection accuracy was calculated via RTG Tools using the vcf produced by the Agilent SureCall Analysis Workflow.

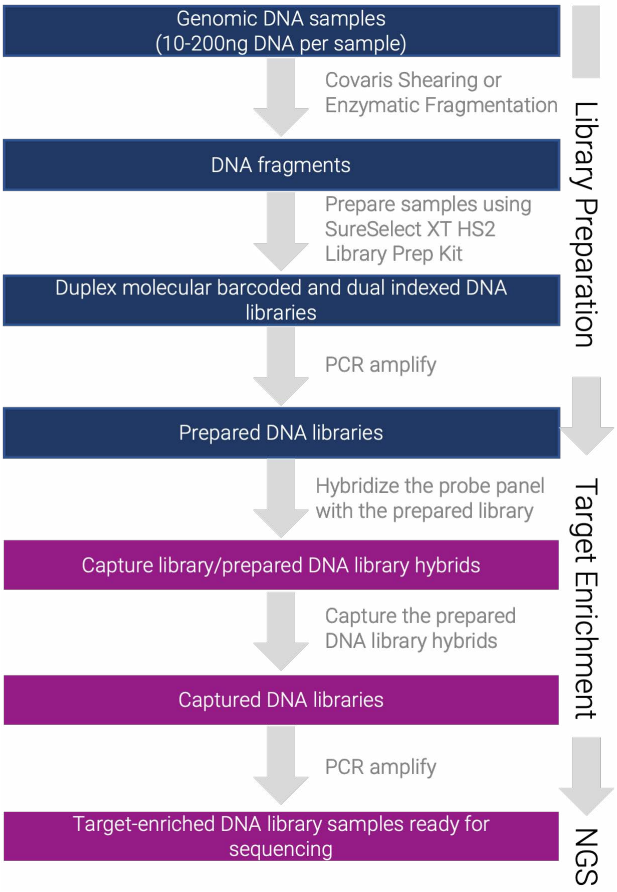

Figure 1 Workflow for SureSelect XT HS2 library preparation comprising adapter ligation, hybridization capture, and PCR.

Results

Table 1 indicates key sequencing metrics, averaged across the two library replicates for each condition. There is a high cross-platform concordance in the percentage of on target reads, percentage of targets having ≥ 30x coverage, and the precision and recall for variant detection. While the G4 Platform and the Illumina HiSeq 4000 both show strong performance, data delivered from the G4 shows a markedly lower rate of optical duplicates for both the custom cancer and XTHS exome v8 panels (0.1% vs 4.0% and 0.6% vs 4.1%, respectively) with a comparable level of PCR duplicates, yielding an overall lower duplicate rate in the G4 data (3.5% vs 8.1% and 5.9% vs 8.9% for the custom cancer and SureSelect Exome v8 panels, respectively). AT and GC coverage values were comparable across platforms **(Table 1 and Figure 3)**, as were precision and recall for variant detection (SNPs and indels). To further examine variant detection performance, we determined the level of overlap for SNPs and indels called with each platform **(Figure 4)**. The majority of SNPs and indels are detected by both platforms, indicating that the data are highly comparable.

Panel	Custom Cancer Panel		SureSelect Exome V8	
Platform	G4	HiSeq 4000	G4	HiSeq 4000
Percent on Target	65.2%	66.5%	71.4%	71.2%
Percent ≥ 30X Coverage	93.3%	92.8%	95.0%	95.0%
Optical Duplicates	0.1%	4.0%	0.6%	4.1%
PCR Duplicates	3.4%	4.1%	5.3%	4.8%
Fold-80	1.69	1.66	1.48	1.45
AT Dropout	1.3	1.6	1.3	2.9
GC Dropout	9.8	6.8	4.4	1.8
Recall	98.8%	98.5%	99.3%	99.4%
Precision	98.2%	98.1%	98.7%	98.9%

Table 1 Key sequencing metrics.

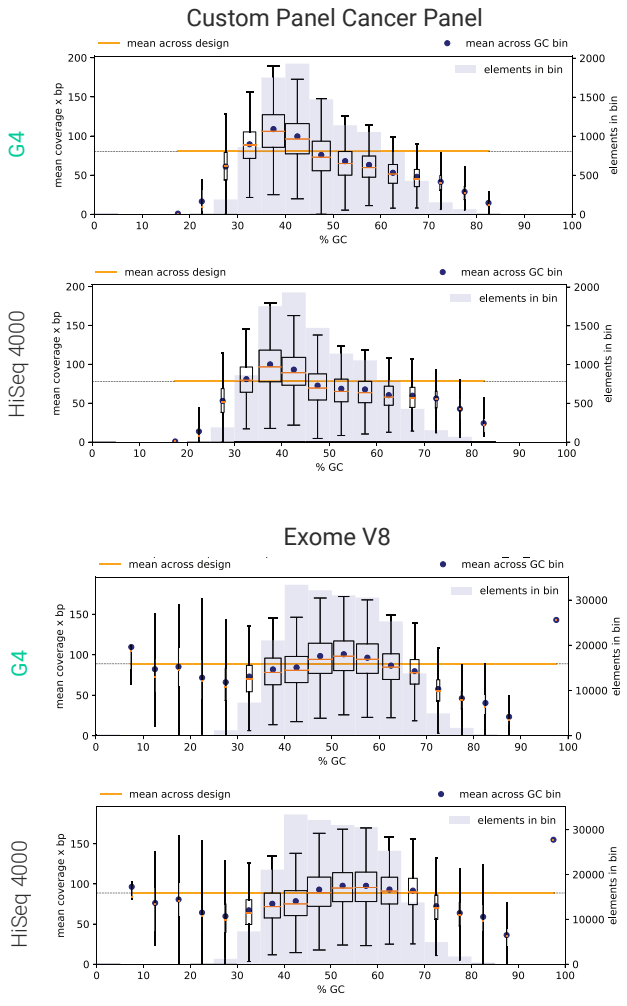

Figure 3 Coverage as a function of GC content. Results reflect the performance of a single representative library of each type.

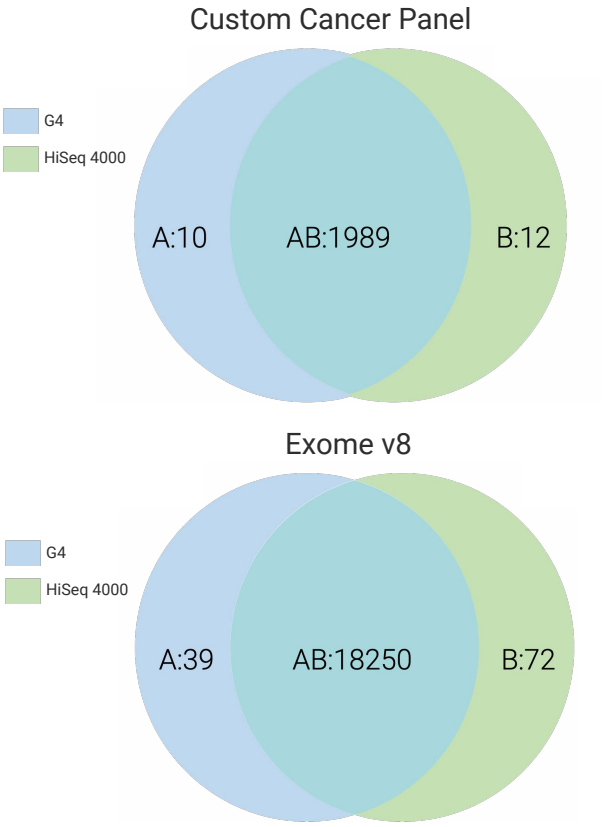

Figure 4 Overlap of variants (SNPs and Indels combined) detected by each platform. Variants were identified via Platypus with default parameters. Results reflect the performance of a single representative library of each type.

Conclusion

Targeted sequencing data generated using the Agilent SureSelect Exome V8 and a custom cancer panel with the G4 Sequencing Platform demonstrate accurate variant analysis. Third-party sequencing libraries produced by Agilent SureSelect kits are easily rendered compatible for analysis on the G4 Platform using the Singular Genomics Compatibility Kit. We observe a high cross-platform overlap in variants detected using the Agilent SureCall Analysis Workflow, consistent with the high accuracy and compatibility of the G4 Platform with existing bioinformatic workflows.

The combination of Agilent SureSelect library preparation and G4 sequencing enables rapid and accurate detection of key variant types for oncology translational and clinical research. The G4 Platform can be seamlessly integrated into labs running targeted sequencing workflows. Labs can benefit from unique flow cell flexibility to match sequencing throughput more precisely to the sample set on hand. This, in combination with shorter sequencing cycle times, enables less waste, reduced turnaround times, and controlled costs for labs incorporating the G4 Sequencing Platform into their Agilent targeted sequencing operations.

SINGULAR GENOMICS

Begin Your Journey with G4

[Contact our sales team](#) to learn more about the capabilites of 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

Contact Agilent Technologies

Learn more about the SureSelect Exome v8 and XT HS2 Kits

Agilent

Website: www.agilent.com
Email: genomicssupport@agilent.com
Address: 5301 Stevens Creek Blvd, Santa Clara, CA 95051