APPLICATION NOTE

Illumina Genomics Architecture enables PopGen studies with Illumina DNA PCR-Free Prep

Consistent library prep performance by automating WGS workflows

illumina®

Introduction

Illumina next-generation sequencing (NGS) technology delivers high-quality, accurate data and enables a broad array of applications in genomics, transcriptomics, and epigenomics. The NGS workflow proceeds from library preparation to sequencing to data analysis and interpretation. The Illumina product portfolio encompasses a range of components throughout the workflow that address the many possible application areas. Because of the diversity of Illumina products, some new customers find that integrating individual components into a single workflow (isolation of genetic material through variant reporting) is laborious and time consuming. Also, while many Illumina workflows are compatible with automation, some customers are daunted by the technical expertise required to integrate and optimize automated methods into their existing practices.

Illumina Genomics Architecture (IGA) addresses these challenges by offering a standardized, modular, and flexible framework for rapidly adopting and implementing automation-compatible DNA-to-answer NGS workflows for both research and clinical research applications. IGA was derived from experience gained in assisting customers implementing DNA-to-data workflows for whole-exome sequencing (WES), whole-genome sequencing (WGS), and Population Genomics (PopGen) programs (Figure 1).

One program benefitting from IGA is the SG100K project. SG100K is a collaborative effort between Illumina and Precision Health Research Singapore (PRECISE) to sequence the genomes of 100,000 healthy individuals in Singapore to better understand Asian genomic diversity. This application note presents internal WGS data generated by independent operators in disparate sites in the United Kingdom and Singapore evaluating scripts within IGA to automate the NGS workflow, as part of the SG100K project.

Methods

IGA features dedicated robot scripts for library preparation and software integrations to automate and streamline sequencing and data analysis.

Library preparation

Sequencing libraries were prepared using Illumina DNA PCR-Free Prep (Illumina, Catalog no. 20041794) from 400 ng of high-quality genomic DNA (gDNA) extracted from blood samples collected from healthy individuals registered as part of the SG100K project. Library prep was automated using the Hamilton STAR liquid-handling platform.

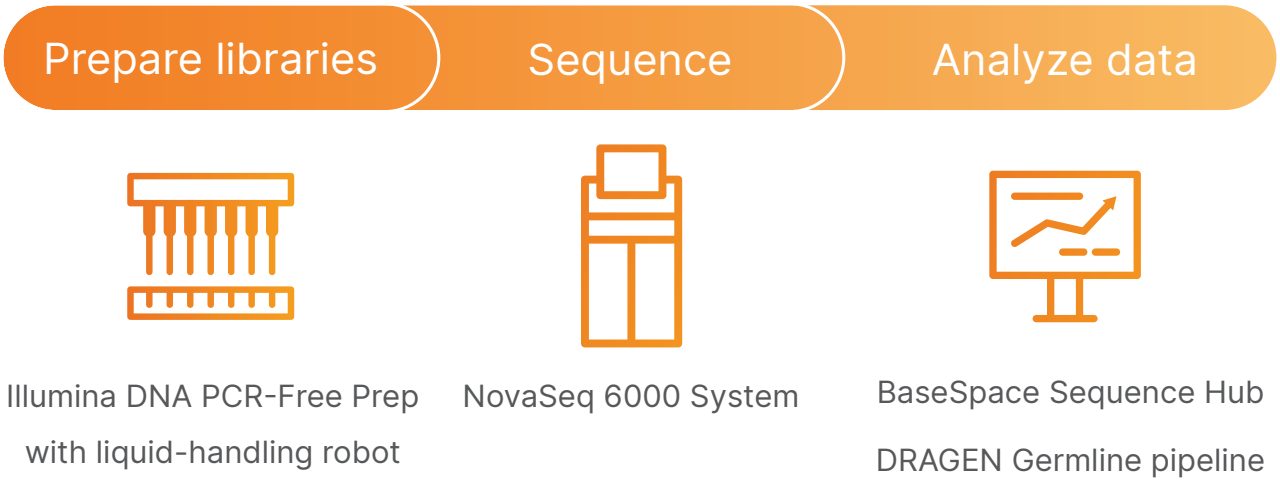

Figure 1: Illumina Genomics Architecture WGS workflow—IGA supports a DNA-to-data workflow for WGS that integrates automated library preparation with Illumina DNA PCR-Free Prep, sequencing on the NovaSeq 6000 System, and analysis with the DRAGEN Germline pipeline.

Sequencing

Prepared libraries were sequenced on the NovaSeq™ 6000 System (Illumina, Catalog no. 20012850) with a run configuration of 2 × 151 bp. Twenty-four samples were run per S4 flow cell at 30× coverage (Illumina, Catalog no. 20028312). As part of the automated workflow with IGA, Clarity™ LIMS software directed the liquid-handling platform to perform bulk pooling, denaturation, and library loading onto the NovaSeq 6000 System and sent the required information to start the sequencing run automatically.

Data analysis

After the sequencing run was complete, data was streamed automatically to BaseSpace™ Sequence Hub for analysis with the DRAGEN™ Germline pipeline v3.7.8. JMP software was used for statistical analysis and plotting graphs.

Results

To evaluate scripts within IGA to automate and streamline a WGS workflow with Illumina DNA PCR-Free Prep, Analysis of Variance (ANOVA) was used to compare sequencing data obtained by independent operators at different sites in Singapore and the United Kingdom. Results demonstrated robust performance and minimal variability across independent operators implementing IGA workflow scripts using separate instrumentation at different sites (Figure 2 and Table 1).

Table 1: Summary of data consistency with IGA

Parameter	Mean value ± standard deviation
Key performance metric (191 samples[a])	
Mean autosomal coverage	36.39 ± 5.9%
Percent of genome with coverage (≥ 15×)	94.72 ± 7.8%
Q30 bases	102.1 ± 16.7 Gbp
Percent autosomal callability	97.05 ± 7.1%
Median insert length	459.4 ± 20.2 bp
Estimated sample contamination	0.001 ± 0.0002
SNV analysis[b] (99 samples)	
SNV recall	99.84 ± 0.11%
SNV precision	99.83 ± 0.03%
Indel recall	99.58 ± 0.31%
Indel precision	99.66 ± 0.15%

a. Extracted DNA from Coriell.
b. Based on data from a single truth sample, NA12878.

Summary

IGA offers a standardized, modular, and flexible framework for rapidly adopting automation-compatible DNA-to-answer NGS workflows. Implementing IGA with Illumina

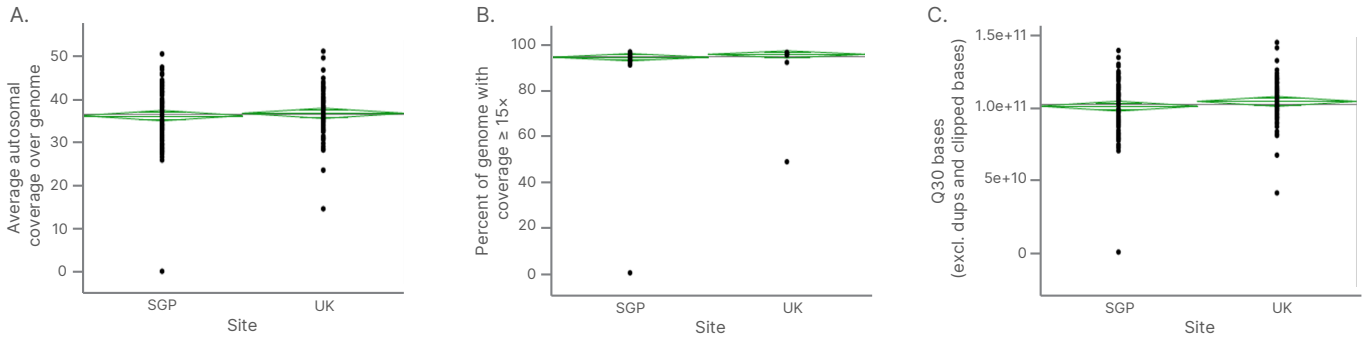

Figure 2: Data performance and consistency with independent operators across different sites—ANOVA of sequencing data across Singapore (SGP) and the United Kingdom (UK) showed highly consistent performance as measured by (A) average autosomal coverage, (B) percent genome with ≥ 15×, and (C) Q30 bases (excluding clipped bases and duplications).

DNA PCR-Free Prep for WGS enables same-day library prep and pooling of 24 samples per S4 flow cell and sequencing to achieve 30× genomic coverage. As part of the PopGen SG100K program, IGA enables highly consistent and robust performance across independent operators at different sites. These results demonstrate the power of IGA to automate workflows and provide reliable performance for NGS methods.

Learn more

Illumina Genomics Architecture

Illumina DNA PCR-Free Prep

NovaSeq 6000 System

DRAGEN secondary analysis

1.800.809.4566 toll-free (US) | +1.858.202.4566 tel
techsupport@illumina.com | www.illumina.com

M-GL-01991 v1.0