Element Biosciences

Data Spotlight

Improved mosaic variant calling and low pass sequencing with Cloudbreak UltraQ™

Introduction

Highly accurate sequencing data is critical for variant detection, particularly in specialized applications like mosaic variant analysis and low-pass whole-genome sequencing (WGS). To improve sequencing accuracy beyond our already high-performing Cloudbreak™ chemistry, we made improvements to the sequencing chemistry, mitigated library prep introduced errors, and thoughtfully handled analysis artifacts for reliable benchmarking. Using these strategies to reduce the most abundant error types, Cloudbreak UltraQ™ sets a new standard for sequencing accuracy and pushes the potential of the AVITI™ system even higher. With the highest accuracy specification on the market for a benchtop sequencer today, UltraQ delivers 70% of bases at Q50 or above (>99.999% accurate) and 90% of bases at Q40 or above (Figure 1).

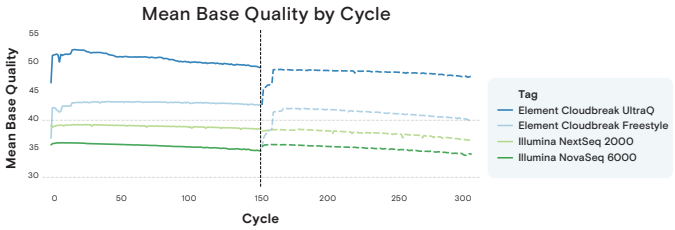
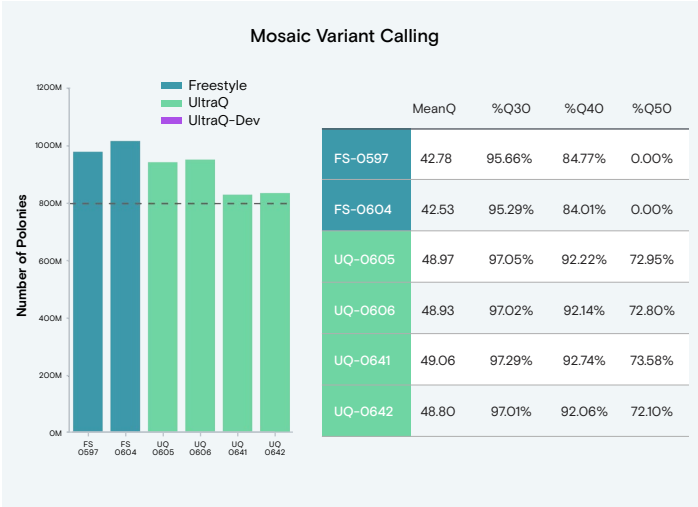


Figure 1. Comparing mean base quality between sequencing kits
In this data spotlight, we demonstrate how sequencing with UltraQ improves performance in mosaic variant calling and low pass WGS applications.

Mosaic Variant Calling

To benchmark mosaic variant calling, we used the mosaic truth set for sample HG002 recently released by the National Institute of Standards and the Medical Device Innovation Consortium (MDIC). The v1.0 truth set contains 85 SNVs and covers nearly 90% of the genome (non-gap bases) so that both sensitivity and precision can be evaluated. PCR-free libraries were generated from NIST Standard Reference Material NA24385 (HG002) using the Kapa Hyper Prep. Two samples were sequenced using our Cloudbreak Freestyle™ (FS) sequencing chemistry and four using Cloudbreak UltraQ (UQ). All libraries were sequenced at 2 x 150 bp yielding an average of ~915M polonies (~831M-1,018M; Figure 2). UltraQ samples demonstrated a higher mean base quality compared to Freestyle samples (Table 2).



	MeanQ	%Q30	%Q40	%Q50
FS-0597	42.78	95.66%	84.77%	0.00%
FS-0604	42.53	95.29%	84.01%	0.00%
UQ-0605	48.97	97.05%	92.22%	72.95%
UQ-0606	48.93	97.02%	92.14%	72.80%
UQ-0641	49.06	97.29%	92.74%	73.58%
UQ-0642	48.80	97.01%	92.06%	72.10%

Figure 2. Average number of polonies and quality metrics across Cloudbreak freestyle and UltraQ sequencing runs

All samples were down sampled to an average of 70x coverage, resulting FASTQ files were aligned to the human genome (hg38 assembly). Variants were called using DeepSomatic v1.7.0 using a tumor-only model and a panel of normal samples (PON), including 20 human WGS samples processed following the same approach and sequenced on AVITI.

All runs show similar recall detecting between 77-83 of the 85 expected SNV mosaic truth sites (Figure 3); however UltraQ runs demonstrated a higher SNV precision, likely driven by the higher quality and lower error reads (Figure 3).

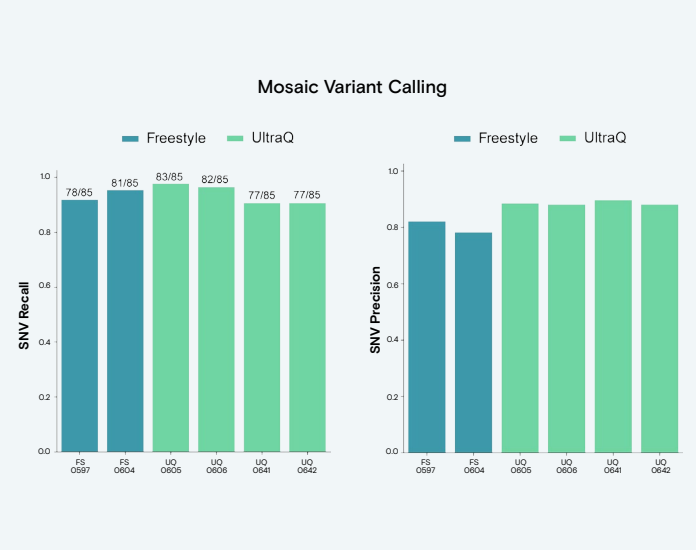


Figure 3. Cloudbreak UltraQ improves SNV precision while maintaining SNV recall.

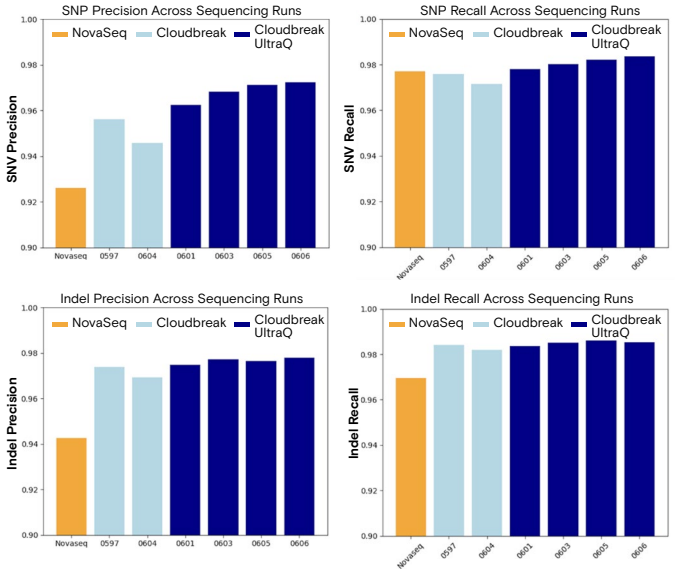


Figure 4. UltraQ improves SNP and indel precision and recall in low pass sequencing analysis.

Low Pass Sequencing Analysis

Low pass sequencing can be used for various applications such as genome wide association studies, ancestry analysis, and agricultural genomics. To determine the impact of UltraQ on sensitivity and precision of small variant calling in low pass sequencing contexts, analysis was performed by the genomics team at Google Research. For this, they compared ten million reads from sample HG002 using Cloudbreak Freestyle and Cloudbreak UltraQ data sequenced on an AVITI and publicly available NovaSeq 6000 data. The sequencing provides roughly 1x coverage of the human genome.

Figure 4 compares sensitivity and precision for SNPs and indels for genomic sites that have evidence for the alternative allele. For SNPs, all data sets show high sensitivity, however, both AVITI data sets show higher precision than the NovaSeq data. Further, the UltraQ data precision is consistently higher than the Cloudbreak Freestyle data, likely owing to the higher accuracy of the assay.

In identifying indels, UltraQ and Cloudbreak Freestyle have similar sensitivity and precision to each other, both outperforming the NovaSeq data (Figure 4).

Conclusion

Cloudbreak UltraQ sequencing delivers >70% of reads at Q50 or higher and 90% of reads at Q40 providing highly accurate sequencing data. Our findings underscore the potential advantages of high-accuracy WGS using Cloudbreak UltraQ for reliable mosaic variant detection and low-pass sequencing applications that can benefit from high data quality.

To learn more, visit elementbiosciences.com