illumina®

Accuracy Improvements in Germline Small Variant Calling with the DRAGEN™ Platform

Several algorithms for accuracy improvements enable small variant detection with high sensitivity and specificity, while also maintaining DRAGEN standards for computation speed

Introduction

With advances in next-generation sequencing (NGS) technology, the volume of generated sequencing data continues to grow exponentially. With this growth comes the demand for fast and efficient analytical methods that maintain high standards in accuracy for variant calling. The Illumina DRAGEN (Dynamic Read Analysis for Genomics) Bio-IT Platform provides highly accurate ultra-rapid secondary analysis of NGS data. The DRAGEN Platform uses highly reconfigurable field-programmable gate array technology (FPGA) to dramatically speed up secondary analysis of NGS data, including mapping, alignment, and variant calling.

Fundamental features of the DRAGEN Platform address common challenges in genomic analysis, such as lengthy compute times and massive volumes of data. The DRAGEN Platform delivers quickness, flexibility, accuracy, and cost effectiveness. The reprogrammable nature of the DRAGEN Platform enables improvement of the algorithms to accommodate new NGS applications. The speed of the platform enables developers to iterate quickly on algorithm designs using computationally intensive methods that are impractical with traditional software only models. As such, the accuracy of the DRAGEN Platform has continuously improved with new versions, and DRAGEN now provides an excellent solution for small variant calling in germline whole-genome sequencing (WGS).

This application note describes recent improvements in the Illumina DRAGEN Bio-IT Platform for rapid secondary analysis, and demonstrates speed and accuracy using three publicly available WGS datasets. We present benchmarking comparisons of DRAGEN v3.2.8 versus other pipelines, including BWA-MEM+GATK4 and DRAGEN v2 (Figure 1). Variant calling results from each pipeline were compared to a "truth set" for reference calls to identify false positives (FPs) and false negatives (FNs). Metrics used for pipeline comparisons are end-to-end run times and accuracy metrics such as recall, precision and FP+FN. The combination of speed, accuracy, and wide range of available applications position the DRAGEN Platform to revolutionize the landscape of genomic analysis.

DRAGEN v3 algorithms for accuracy improvements

DRAGEN v3 implements the latest algorithm updates for detection of single-nucleotide polymorphisms (SNPs) and insertions/deletions (indels), which provide improvements in the precision and analytical sensitivity. Improvements were made in four areas for variant calling: Sample-specific indel error model, rigorous mathematical models of correlated pileup errors, an optimized approach to exhaustively represent an exponential number of haplotype candidates in variant-rich or noisy regions, and column-wise augmentation of the list of events generated by De-Bruijn graph assembly. These upgrades result in modest speed up gains, while elevating standards in accuracy compared to pipelines evaluated in this paper. Each algorithm improvement is described in more detail in the appendix.

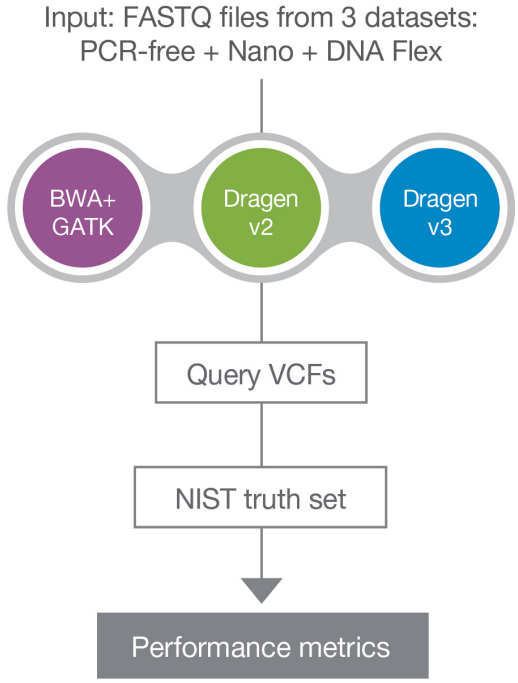

Figure 1: Design of benchmarking comparison study—FASTQ files from 3 datasets were run through three analysis pipelines to generate query VCF files. The Variant Calling Assessment Tool (VCAT) was then used to identify TPs, FPs, and FNs, based on comparison of variant calls to reference variants in the NIST truth set.

Methods

Recommended best practices for benchmarking were followed closely.[1] To demonstrate speed and accuracy with DRAGEN v3, a comparison study was done using three datasets, from different library preparations, generated from the NA12878 sample (Figure 1). Briefly, the FASTQ file from each dataset was used as input for secondary analysis from independent pipelines (DRAGEN v3.2.8, DRAGEN v2, and BWA+GATK[2]). Resultant VCF files from each pipeline (QUERY VCFs) were uploaded to a project in BaseSpace™ Sequence Hub. The Variant Calling Assessment Tool (VCAT v3.1.1 with Hap.py version 0.3.10) was used to compare each QUERY VCF file to a reference variant “truth set” in order to identify true or false variant calls. Results were collected and tabulated for comparisons between pipelines. All input data, analysis results, and evaluation tools are freely available in the BaseSpace project.[3] More detailed descriptions of methods are described in the Appendix.

Results of benchmarking comparisons

Results for both run times and accuracy comparisons demonstrate that DRAGEN provides a powerful solution for secondary analysis of NGS data.

DRAGEN accuracy: FP+FN, recall, and precision

Although DRAGEN v2 was already competitive with industry-leading informatics solution, DRAGEN v3 has several new modifications (described in the algorithm methods section) that result in significant accuracy improvements. The results of this benchmarking comparison also demonstrate that DRAGEN v3 improvements make it superior compared to other popular analysis pipelines, including a previous version of DRAGEN, for all accuracy metrics analyzed in the study.

When the FP+FN metric was evaluated for SNV detection, DRAGEN v3 performed with significantly higher accuracy than both the BWA+GATK4 and DRAGEN v2 pipeline for all three datasets (Figure 2). When the FP+FN metric was evaluated for indel detection, DRAGEN v3 performed better than the BWA+GATK4 pipeline for all three datasets, while also showing further improvement between DRAGEN v3 and DRAGEN v2 (Figure 3).

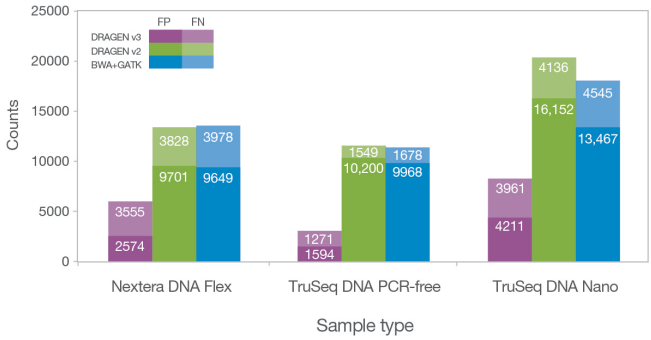

Figure 2: False positives and false negatives with SNV detection—Raw data files (FASTQ) from three datasets were analyzed by three independent pipelines. Each dataset (TruSeq DNA PCR-free, Nextera DNA Flex, and TruSeq DNA Nano) was generated from NA12878 sample DNA, and variant calls (VCF) from each analysis pipeline were compared to NIST truth set (also based on NA12878 sample) to identify FPs and FNs.

When evaluating the precision and recall metrics, the advantage of DRAGEN v3 algorithm improvements are evident for both SNP and indel detection. Values for both precision and recall are consistently above 99% for all pipelines and with each SNV detection dataset (Table 1). For SNP detection, DRAGEN v2 was comparable with BWA+GATK4. But DRAGEN v3 shows significant improvement in both recall and precision over the other two pipelines. For indel detection, DRAGEN v2 showed higher accuracy than BWA+GATK4, while DRAGEN v3 yielded further improvement over DRAGEN v2 for both recall and precision (Table 2).

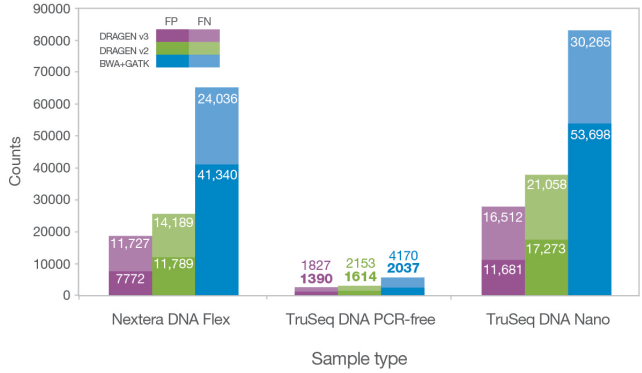

Figure 3: False positives and false negatives with indel detection—Raw data files (FASTQ) from three datasets were analyzed by three independent pipelines. Each dataset (TruSeq DNA PCR-free, Nextera DNA Flex, and TruSeq DNA Nano) was generated from NA12878 sample DNA, and variant calls (VCF) from each analysis pipeline were compared to NIST truth set (also based on NA12878 sample DNA) to identify FPs and FNs.

Table 1: Sensitivity and specificity of SNV detection

	Precision			Recall		
Datasets	DRAGEN v3	DRAGEN v2	BWA+GATK	DRAGEN v3	DRAGEN v2	BWA+GATK
TruSeq DNA PCR-free	99.95%	99.68%	99.69%	99.96%	99.95%	99.95%
Nextera DNA Flex	99.92%	99.70%	99.70%	99.89%	99.88%	99.88%
TruSeq DNA Nano	99.87%	99.50%	99.58%	99.88%	99.87%	99.86%

Table 2: Sensitivity and specificity of indel detection

	Precision			Recall		
Datasets	DRAGEN v3	DRAGEN v2	BWA+GATK	DRAGEN v3	DRAGEN v2	BWA+GATK
TruSeq DNA PCR-free	99.71%	99.66%	99.58%	99.62%	99.55%	99.13%
Nextera DNA Flex	98.37%	97.54%	91.53%	97.56%	97.05%	95.01%
TruSeq DNA Nano	97.56%	96.39%	89.37%	96.57%	95.63%	93.71%

DRAGEN speed

DRAGEN run times comparison were collected for both the on-cloud and on-site solutions. For the on-site solution, DRAGEN v3 was compared to BWA+GATK with both pipelines run on the same server. For the on-cloud solution, DRAGEN v3 run on BaseSpace Sequence Hub was compared to BWA+GATK run on Terra.[4]

DRAGEN accelerates both the mapping process and variant calling, which can be run independently. Though not captured here, it is also worth noting that upstream of the secondary analysis, DRAGEN also supports accelerated BCL2FASTQ conversion, which greatly improves speed and efficiency while producing identical FASTQs. Also worth noting is that DRAGEN automatically outputs an exhaustive list of QC metrics, both at the mapping and variant calling level, with little to no run time overhead. This is in contrast to other pipelines which rely on slow-running third party tools (eg, Samtools, Picard) to acquire QC metrics with significant run time overhead.

When executions speeds were measured with pipelines run on the same on-premise server, DRAGEN v3 was significantly faster than BWA+GATK, with speed up gains in the range of 16-18× (Figure 4).

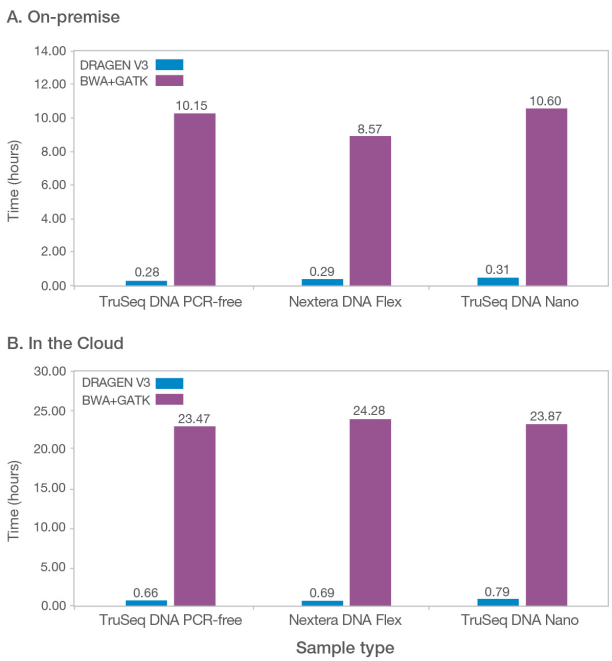

Figure 4: Analysis run time comparisons on-premise and in the cloud—(A) DRAGEN v3 and BWA+GATK run on the same on-premise server. (B) DRAGEN v3 run on BaseSpace Sequence Hub was compared to BWA+GATK run on Terra.

When executions speeds were measured with pipelines run in the cloud, DRAGEN v3 on BaseSpace Sequence Hub was significantly faster than BWA+GATK run on Terra, with speed up gains in the range of 13-16×.

Summary

As genomic applications move towards precise characterization of difficult regions of the genome, and measuring low allele frequency calls from samples with high noise level, DRAGEN proves to be the best-suited platform to process the NGS data of the future both efficiently and accurately.

DRAGEN speed not only enables researchers to keep up with the increasing throughput of NGS instruments, but just as importantly, it also enables fast iterations for continuous improvement of its algorithms to provide high accuracy.

Appendix

Detailed description of new algorithms

Sample-specific PCR error model

One of the challenges in variant calling is distinguishing indel errors from true variants. To do so, variant callers often employ a Hidden Markov Model (HMM), which models the statistical behavior of indel errors, as part of the probability calculation. The HMM typically has input parameters, Gap Open Penalty (GOP) and Gap Continuation Penalty (GCP), which are directly related to the indel error rate (ie, indel error rate = f(GOP,GCP)). Indel errors are more likely in the presence of short tandem repeats (STRs), and the error probability (and thus GOP and GCP) may depend on both the period and the length of the STR. The error process may differ significantly from one dataset to another, depending on factors such as PCR amplification. For accurate detection, it is important to use HMM parameters that accurately model the error process on a per sample basis. However, typical variant callers often use fixed parameters or non-sample-specific predetermined functions that fail to accurately model the error process, resulting in poor detection performance.

The HMM auto calibration implemented in DRAGEN v3 addresses the above problem by estimating the PCR parameters directly from the dataset being processed. This operation is performed after mapping & alignment and prior to variant calling, without knowledge of the ground truth and without using external databases of known mutations. The parameters depend on both the STR period and the repeat length.

For a given STR period and length, a set of N loci with the desired period and length is selected, and the algorithm examines the pileups of reads mapped to those loci, counting the indels observed at each locus. The key idea is that by considering a sufficient number of loci, it's possible to accurately estimate the parameters of interest. We do so by finding the parameters that maximize the probability of producing the set of N observed pileups. If the number of parameters to maximize the probability over is small enough (eg, 2), an exhaustive search is possible. In the current implementation of DRAGEN v3, the optimization is performed over two parameters: GOP and alpha, which indicates the probability of indel variants of any non-zero length. For each STR period and length considered, the search outputs GOP and

alpha that maximize the probability of producing the set of N observed pileups, and those values are used as input to the HMM. Extending the search beyond 2 parameters is also possible and would provide further improvements.

Base quality dropoff (BQD)

Conventional variant callers are designed with the assumption that sequencing errors are independent across reads; following this assumption, it's very unlikely that multiple identical errors will occur at a specific locus. However, after analyzing NGS datasets, it was observed that bursts of errors are far more common than would be predicted by the independence assumption, and these bursts can result in lots of false positives.

Fortunately, these errors have distinct characteristics differentiating them from true variants. The BQD (base quality drop off) algorithm implemented in DRAGEN v3 is a detection mechanism that exploits certain properties of those errors (strand bias, localization of the error in the read, low mean base quality, at the locus of interest) and incorporates them into the probability calculation in a simple and robust manner, in the genotyper. New genotype candidates hypotheses are added to the legacy list of diploid genotypes (those that assume independent pileup errors). For example, in the case of a locus with 1 ALT allele, in addition to considering P(G00|R), P(G01|R), P(G11|R), we add two more hypotheses as P(G00,E1|R) and P(G11,E0|R), where allele E0 and E1 represent reference allele and ALT allele coming from a sequencing error. The properties of those errors, such as strand bias, localization of the error in the read and mean base quality are incorporated in the calculation of P(G00,E1|R) and P(G11,E0|R). Then the winning genotype is taken over max (max(P(G00|R), P (G00,E1|R)), P(G01|R), max(P(G11|R), P(G11,E0|R))).

Being able to characterize correlated sequencing errors from within the core of the variant caller results in a significant gain in specificity because a lot of FP calls are removed. It also helps sensitivity by correcting genotype errors.

Foreign read detection (FRD)

Conventional variant callers treat mapping errors as independent error events per read, ignoring the fact that such errors typically occur in bursts. This can result in variant calls emitted with very high confidence scores in spite of low MAPQ and/or skewed AF. To mitigate this problem, conventional variant callers typically filter out reads upstream of variant calling, based on a MAPQ threshold (i.e., reads with MAPQ< threshold are excluded from the calculation). However, this discards valuable evidence from within the variant caller and does a poor job of suppressing false positives.

DRAGEN v3 has implemented Foreign Read Detection (FRD), which is an extension to the legacy genotyping algorithm by incorporating an additional hypothesis that some read(s) in the pileup are foreign reads (i.e., their true location is elsewhere in the reference genome and/or are originated from outside of the reference genome (i.e. sample contamination)). The algorithm exploits multiple properties (skewed allele frequency and low MAPQ) and incorporates this evidence into the probability calculation in a mathematically rigorous manner.

New genotype candidates hypotheses are added to the legacy list of diploid genotypes (those that assume independent pileup errors). For example, in the case of a locus with 1 ALT allele, in addition to considering P(G00|R), P(G01|R), P(G11|R), we add two more hypotheses as P(G00,F1|R) and P(G11,F0|R), where allele F0 and F1 represent reference allele and ALT allele coming from a mapping error. The properties of those errors, such as allele depth and MAPQ are incorporated in the calculation of P(G00,F1|R) and P(G11,F0|R). Then the winning genotype is taken over max (max(P (G00|R), P(G00,F1|R)), P(G01|R), max(P(G11|R), P(G11,F0|R))).

Sensitivity is improved from rescuing FN, correcting genotypes and enabling lowering of the MAPQ threshold for incoming reads into the variant caller. Specificity is improved from removing FP and correcting genotypes.

FRD is a more powerful tool than post-VCF filtering approaches to improve F-measure, because, rather than simply detecting suspicious results (e.g. based on allele depth or read errors) post variant caller, the detection algorithm directly incorporates the presence of foreign reads via rigorous maximum-likelihood detection.

PDHMM and column-wise detection

Variant callers, such as GATK Haplotype Caller and DRAGEN,use Debruijn Graph to re-assemble reads in order to determine candidate haplotypes and identify potential variant sites. In regions of the genome with tandem repeats, structural variants, or clusters of sequencing errors, lower sensitivity can result from failure of the graph assembly methodology to give a complete list of candidate haplotypes and variant sites.

Column-wise event detection supplements the Debruijn graph by scanning each column of an active region for potential variant sites (SNPs and indels) and completing the list of candidate haplotypes. This restores sensitivity in regions where the graph fails.

Impact of FRD/BQD on QUAL/GQ/QD and post-VCF hard-filtering

DRAGEN v3 Variant Caller has implemented two algorithms that model correlated errors across reads in a given pileup, foreign read detection (FRD) to detect mismapped reads, and base quality drop off (BQD) algorithm to detect correlated base call errors. Besides improving specificity and sensitivity, these two algorithms have an impact/benefit at two levels:

Confidence score (QUAL, GQ, QD) values are in a realistic Phred-scale range.

Conventional variant callers typically output inflated QUAL values in Phred scale in the range of few thousands which have no practical meaning statistically. Modeling correlated errors from within the variant caller brings back these values to a statistically realistic and meaningful range.

Dependency on post VCF filtering rules is substantially reduced.

In conventional variant callers, because of the variant caller's inability to distinguish between correlated errors and true variants, it was necessary to apply hard-filtering rules post VCF to filter out the excess number of FP calls. Several VCF annotations (e.g., QD, MQ, FS, MQRankSum) were compared to ad-hoc thresholds, to flag calls as FP. Alternatively, those annotations could be fed to a machine learning algorithm and trained against a truth set, and false positive could then be filtered out based on the training (e.g., VQSR).

In DRAGEN v3, the algorithms were improved at the core of the variant caller and therefore the dependency on post VCF filtering was substantially reduced. DRAGEN v3 default hard-filtering rule simply uses QUAL with a threshold corresponding to the best Fmeas (best tradeoff between sensitivity and specificity).

Detailed methods

Input datasets

Three data sets were selected to represent multiple library preparation methods, both including and excluding PCR (TruSeq DNA Nano, TruSeq DNA PCR-Free, and Nextera DNA Flex). Each data set was generated using NA12878 sample DNA. Following DNA library preparation according to respective reference guides,[5-7] resulting libraries were sequenced in 2×150 paired end runs on the NovaSeq™ 6000 System. To normalize the number of reads, each data set was downsampled to 30× coverage with the FASTQ Toolkit in BaseSpace Sequence Hub. All three datasets are publicly available at BaseSpace Sequence Hub, so that independent assessment of results can be performed.

Human reference genome

The genome reference used was Human hs37d5 in the DRAGEN BaseSpace app, and the equivalent genome reference was used in local analysis for each pipeline under evaluation. This reference includes decoys.[8]

Secondary analysis pipelines

We compare three secondary analysis pipelines. The first pipeline is DRAGEN v2 end to end (DRAGEN used for both mapping and alignment stage and variant calling). The second pipeline is DRAGEN v3 end to end. The third pipeline uses BWA-MEM for the mapping and alignment stage and GATK4-HC for the variant calling stage.

To make a fair comparison, we applied the same hard filtering rule for all three pipelines, which consisted of applying a GQ threshold to the prefilter VCFs. The threshold was selected to be close to the best Fmeas point for each pipeline (Table 3).

Table 3: Optimal Fmeas QC thresholds

GQ for best Fmeas	SNP	Indel
DRAGEN v3.2.8	9	9
DRAGEN v2.5	2	8
BWA+GATK	1	2

DRAGEN was run on an on-premise server, as well as in the cloud using BaseSpace Sequence Hub. Although computing time is slightly longer in the cloud, the variant calling results do not differ. The BWA+GATK pipeline was run on the same on-premise server as DRAGEN, where the BCBIO framework was installed.[9] BCBIO runs the BWA+GATK following the GATK best practice guidelines, and also applies additional optimizations to improve parallelism for run time speed up. For on-cloud analysis, the BWA+GATK pipeline was run on Terra.

DRAGEN 3.3.0

DRAGEN App version:
DRAGEN Germline Pipeline 3.2.8
DRAGEN Host Software Version 05.011.281.3.2.8

BWA-Mem (0.7.17) + GATK4 (4.0.2)

Table 4: Parameters from the configuration file of BCBIO algorithms

Parameter	Value
align_split_size	5000000
aligner	BWA
coverage_depth	High
coverage_interval	Regional
mark_duplicates	True
merge_bamprep	False
platform	Illumina
quality_format	Standard
realign	False
recalibrate	False
tools_off	Vqsr
variantcaller	GATK-haplotype

analysis: variant2
resources: gatk-haplotype

BWA+GATK on Terra

Analysis ready Bam files from BWA-Mem (from BCBIO runs) were used as inputs to run GATK on Terra. Briefly, we followed the GATK4-germline-snps-indels (https://github.com/gatk-workflows/gatk4-germline-snps-indels) workflow with modifications in specific parameters to match the parameters in BCBIO runs. All the runs were executed with a free trial account on Terra.

Exact WDL method is available in BaseSpace Sequence Hub public data.

WDL Method configurations:
GATK docker image: broadinstitute/gatk:4.0.2.0
GITC docker: broadinstitute/genomes-in-the-cloud:2.3.1-1500064817

Reference fasta: hs37d5 (same as other pipelines)

Table 6: Definitions and calculations for metrics involved with precision and recall

Metric	Common name	Definition	Formula
TRUTH.TP	True positives (Truth)	Number of truth calls for which there is a query call that is consistent with the truth call and its genotype	
QUERY.TP	True positives (Query)	Number of query calls for which there is a truth call that is consistent with the query call and its genotype.	
TRUTH.FN	False negatives	Number of truth calls for which there is no query call that is consistent with the truth call and its genotype	
QUERY.FP	False positives	Number of query calls for which there is no truth call that is consistent with the query call and its genotype.	
METRIC.Recall	Recall, sensitivity	Fraction of truth calls that are consistent with a query allele and genotype call within the confident regions	TRUTH.TP / (TRUTH.TP + TRUTH.FN)
METRIC.Precision	Precision, positive predictive value	Fraction of query calls that are consistent with a truth allele and genotype call within the confident regions	QUERY.TP / (QUERY.TP + QUERY.FP)

Only raw VCF files were generated in this pipeline. Post filtering was performed locally. Raw VCF files are available in BaseSpace Sequence Hub public data.

Basespace (January 2019) Specify the spec of the AWS F1 instance used (AWS F1 4× large).
BaseSpace Sequence Hub app version: 3.2.8

Table 5: Local Server (CentOS 7 x86_64, Supermicro 1029)

Part	Full Model Name	Notes
Chassis	SYS-1029GQ-TNRT	1 rack unit
CPU	2 x Intel(R) Xeon(R) Gold 6126 CPU @ 2.60GHz	24 cores, 48 threads
RAM	384GB	DDR4, 2666 MHz
Staging	Intel SSDPE2KE020T7	2TB NVME

Benchmark truth set (NIST)

Benchmarking of variant calls requires a specific reference genome and an associated set of calls that represent the "true answers" for that genome. Such call sets have the property that they can be used as "truth" to accurately identify false positives and negatives. For this study, the used truth set was based on reference calls based on the same source of DNA (NA12878) that were established by National Institute of Standards and Technology (NIST). The Genome in a Bottle Consortium (GIAB) is a public-private-academic consortium hosted by NIST. GIAB published a benchmark set of small variant and reference calls for its pilot genome, NA12878, characterizing a high-confidence genotype for approximately 90% of GRCh37 and GRCh38.

True positives (TPs) are variant calls that agree with reference calls from the NIST truth set. False positives (FPs) are variant calls that do not exist in the truth set, and false negatives (FNs) are variants in the truth set that were not called in the QUERY VCF.

The Variant Calling Assessment Tool (VCAT) was used to compare each QUERY VCF file to the NIST truth set v3.3.2. This tool runs hap.py using the RTG vcfeval evaluation engine. TPs, FPs, and FNs were determined by hap.py's output files *roc.Locations.INDEL.csv and *roc.Locations.SNP.csv of TRUTH TP, QUERY.TP, QUERY.FP, and TRUTH.FN.

The matching stringency type used to calculate TP, FP and FN is "genotype match" (cf. [1]), for which only sites with matching alleles and genotypes are counted as TP. This means genotype errors and allele mismatches are counted as both FPs and FNs.

Benchmarking evaluation metrics

For speed comparisons, the total run time in seconds, from FASTQ to VCF, is derived from analysis log files and/or from analysis times shown in reports.

To perform accuracy comparisons across various pipelines, we use recommended standards in performance metrics (Table 6).[1] Precision is the metric representing analytical specificity, or the ability to correctly identify the absence of variants or "absence of false positives". Recall is the metric representing analytical sensitivity, or the ability to detect variants that are known to be present or "absence of false negatives".

Definitions and calculations for metrics involved with the precision and recall numbers are as based on reference.

References

1. Krusche P, Trigg L, Boutros PC, et al. Best practices for benchmarking germline small-variant calls in human genomes. *Nat Biotechnol*. 2019;37(5):555-560.
2. GATK Best Practices. software.broadinstitute.org/gatk/best-practices/. Accessed May 9, 2019.
3. The BaseSpace project. basespace.illumina.com/s/3ExEZMlH8Lkq. Accessed May 15, 2019.
4. FireCloud Powered by Terra. firecloud.terra.bio/. Accessed May 15, 2019.
5. Illumina (2017) TruSeq DNA PCR-Free Reference Guide. Accessed March 6. 2019.
6. Illumina (2017) TruSeq DNA Nano Reference Guide. Accessed March 6. 2019.
7. Illumina (2018) Nextera DNA Flex Library Prep Reference Guide. Accessed March 6. 2019.
8. hs37d5 Reference Genome. ftp-trace.ncbi.nih.gov/1000genomes/ftp/technical/reference/phase2_reference_assembly_sequence/. Accessed May 9, 2019.
9. Bcbio-nextgen. Docs. bcbio-nextgen.readthedocs.io/en/latest/. Accessed May 9, 2019.

Illumina, Inc. • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com