USAGE: <program name> [-h]

Available Programs:

--------------------------------------------------------------------------------------

Base Calling: Tools that process sequencing machine data, e.g. Illumina base calls, and detect sequencing level attributes, e.g. adapters

CheckIlluminaDirectory (Picard) Asserts the validity for specified Illumina basecalling data.

CollectIlluminaBasecallingMetrics (Picard) Collects Illumina Basecalling metrics for a sequencing run.

CollectIlluminaLaneMetrics (Picard) Collects Illumina lane metrics for the given BaseCalling analysis directory.

ExtractIlluminaBarcodes (Picard) Tool determines the barcode for each read in an Illumina lane.

IlluminaBasecallsToFastq (Picard) Generate FASTQ file(s) from Illumina basecall read data.

IlluminaBasecallsToSam (Picard) Transforms raw Illumina sequencing data into an unmapped SAM or BAM file.

MarkIlluminaAdapters (Picard) Reads a SAM or BAM file and rewrites it with new adapter-trimming tags.

--------------------------------------------------------------------------------------

Copy Number Variant Discovery: Tools that analyze read coverage to detect copy number variants.

AnnotateIntervals (BETA Tool) Annotates intervals with GC content

CallCopyRatioSegments (BETA Tool) Calls copy-ratio segments as amplified, deleted, or copy-number neutral

CombineSegmentBreakpoints (EXPERIMENTAL Tool) Combine the breakpoints of two segment files and annotate the resulting intervals with chosen columns from each file.

CreateReadCountPanelOfNormals (BETA Tool) Creates a panel of normals for read-count denoising

DenoiseReadCounts (BETA Tool) Denoises read counts to produce denoised copy ratios

DetermineGermlineContigPloidy (BETA Tool) Determines the baseline contig ploidy for germline samples given counts data

GermlineCNVCaller (BETA Tool) Calls copy-number variants in germline samples given their counts and the output of DetermineGermlineContigPloidy

MergeAnnotatedRegions (EXPERIMENTAL Tool) (EXPERIMENTAL) Merge annotated genomic regions based entirely on contig and annotation value.

ModelSegments (BETA Tool) Models segmented copy ratios from denoised read counts and segmented minor-allele fractions from allelic counts

PlotDenoisedCopyRatios (BETA Tool) Creates plots of denoised copy ratios

PlotModeledSegments (BETA Tool) Creates plots of denoised and segmented copy-ratio and minor-allele-fraction estimates

PostprocessGermlineCNVCalls (BETA Tool) Postprocesses the output of GermlineCNVCaller and generates VCF files.

TagGermlineEvents (EXPERIMENTAL Tool) (EXPERIMENTAL) Do a simplistic tagging of germline events in a tumor segment file.

--------------------------------------------------------------------------------------

Coverage Analysis: Tools that count coverage, e.g. depth per allele

ASEReadCounter Generates table of filtered base counts at het sites for allele specific expression

CollectAllelicCounts (BETA Tool) Collects reference and alternate allele counts at specified sites

CollectReadCounts (BETA Tool) Collects read counts at specified intervals

CountBases Count bases in a SAM/BAM/CRAM file

CountBasesSpark (BETA Tool) Counts bases in the input SAM/BAM

CountReads Count reads in a SAM/BAM/CRAM file

CountReadsSpark (BETA Tool) Counts reads in the input SAM/BAM

GetPileupSummaries (BETA Tool) Tabulates pileup metrics for inferring contamination

Pileup Prints read alignments in samtools pileup format

PileupSpark (BETA Tool) Prints read alignments in samtools pileup format

--------------------------------------------------------------------------------------

Diagnostics and Quality Control: Tools that collect sequencing quality related and comparative metrics

AccumulateVariantCallingMetrics (Picard) Combines multiple Variant Calling Metrics files into a single file

AnalyzeCovariates Evaluate and compare base quality score recalibration (BQSR) tables

BamIndexStats (Picard) Generate index statistics from a BAM file

CalcMetadataSpark (BETA Tool) (Internal) Collects read metrics relevant to structural variant discovery

CalculateContamination Calculate the fraction of reads coming from cross-sample contamination

CalculateReadGroupChecksum (Picard) Creates a hash code based on the read groups (RG).

CheckFingerprint (Picard) Computes a fingerprint from the supplied input (SAM/BAM or VCF) file and compares it to the provided genotypes

CheckPileup Compare GATK's internal pileup to a reference Samtools mpileup

CheckTerminatorBlock (Picard) Asserts the provided gzip file's (e.g., BAM) last block is well-formed; RC 100 otherwise

ClusterCrosscheckMetrics (Picard) Clusters the results of a CrosscheckFingerprints run by LOD score

CollectAlignmentSummaryMetrics (Picard) <b>Produces a summary of alignment metrics from a SAM or BAM file.</b>

CollectBaseDistributionByCycle (Picard) Chart the nucleotide distribution per cycle in a SAM or BAM file

CollectBaseDistributionByCycleSpark (BETA Tool) Collects base distribution per cycle in SAM/BAM/CRAM file(s).

CollectGcBiasMetrics (Picard) Collect metrics regarding GC bias.

CollectHiSeqXPfFailMetrics (Picard) Classify PF-Failing reads in a HiSeqX Illumina Basecalling directory into various categories.

CollectHsMetrics (Picard) Collects hybrid-selection (HS) metrics for a SAM or BAM file.

CollectIndependentReplicateMetrics (Picard) (EXPERIMENTAL Tool) Estimates the rate of independent replication of reads within a bam.

CollectInsertSizeMetrics (Picard) Collect metrics about the insert size distribution of a paired-end library.

CollectInsertSizeMetricsSpark (BETA Tool) Collects insert size distribution information on alignment data

CollectJumpingLibraryMetrics (Picard) Collect jumping library metrics.

CollectMultipleMetrics (Picard) Collect multiple classes of metrics.

CollectMultipleMetricsSpark (BETA Tool) Runs multiple metrics collection modules for a given alignment file

CollectOxoGMetrics (Picard) Collect metrics to assess oxidative artifacts.

CollectQualityYieldMetrics (Picard) Collect metrics about reads that pass quality thresholds and Illumina-specific filters.

CollectQualityYieldMetricsSpark (BETA Tool) Collects quality yield metrics from SAM/BAM/CRAM file(s).

CollectRawWgsMetrics (Picard) Collect whole genome sequencing-related metrics.

CollectRnaSeqMetrics (Picard) Produces RNA alignment metrics for a SAM or BAM file.

CollectRrbsMetrics (Picard) <b>Collects metrics from reduced representation bisulfite sequencing (Rrbs) data.</b>

CollectSequencingArtifactMetrics (Picard) Collect metrics to quantify single-base sequencing artifacts.

CollectTargetedPcrMetrics (Picard) Calculate PCR-related metrics from targeted sequencing data.

CollectVariantCallingMetrics (Picard) Collects per-sample and aggregate (spanning all samples) metrics from the provided VCF file

CollectWgsMetrics (Picard) Collect metrics about coverage and performance of whole genome sequencing (WGS) experiments.

CollectWgsMetricsWithNonZeroCoverage (Picard) (EXPERIMENTAL Tool) Collect metrics about coverage and performance of whole genome sequencing (WGS) experiments.

CompareBaseQualities Compares the base qualities of two SAM/BAM/CRAM files

CompareDuplicatesSpark (BETA Tool) Determine if two potentially identical BAMs have the same duplicate reads

CompareMetrics (Picard) Compare two metrics files.

CompareSAMs (Picard) Compare two input ".sam" or ".bam" files.

ConvertSequencingArtifactToOxoG (Picard) Extract OxoG metrics from generalized artifacts metrics.

CrosscheckFingerprints (Picard) Checks that all data in the input files appear to have come from the same individual

CrosscheckReadGroupFingerprints (Picard) DEPRECATED: USE CrosscheckFingerprints. Checks if all read groups appear to come from the same individual.

EstimateLibraryComplexity (Picard) Estimates the numbers of unique molecules in a sequencing library.

EstimateLibraryComplexityGATK (BETA Tool) Estimate library complexity from the sequence of read pairs

FlagStat Accumulate flag statistics given a BAM file

FlagStatSpark (BETA Tool) Spark tool to accumulate flag statistics

GetSampleName (BETA Tool) Emit a single sample name

IdentifyContaminant (Picard) Computes a fingerprint from the supplied SAM/BAM file, given a contamination estimate.

MeanQualityByCycle (Picard) Collect mean quality by cycle.

MeanQualityByCycleSpark (BETA Tool) MeanQualityByCycle on Spark

QualityScoreDistribution (Picard) Chart the distribution of quality scores.

QualityScoreDistributionSpark (BETA Tool) QualityScoreDistribution on Spark

ValidateSamFile (Picard) Validates a SAM or BAM file.

ViewSam (Picard) Prints a SAM or BAM file to the screen

--------------------------------------------------------------------------------------

Intervals Manipulation: Tools that process genomic intervals in various formats

BedToIntervalList (Picard) Converts a BED file to a Picard Interval List.

IntervalListToBed (Picard) Converts an Picard IntervalList file to a BED file.

IntervalListTools (Picard) A tool for performing various IntervalList manipulations

LiftOverIntervalList (Picard) Lifts over an interval list from one reference build to another.

PreprocessIntervals (BETA Tool) Prepares bins for coverage collection

SplitIntervals Split intervals into sub-interval files.

--------------------------------------------------------------------------------------

Metagenomics: Tools that perform metagenomic analysis, e.g. microbial community composition and pathogen detection

PathSeqBuildKmers Builds set of host reference k-mers

PathSeqBuildReferenceTaxonomy Builds a taxonomy datafile of the microbe reference

PathSeqBwaSpark Step 2: Aligns reads to the microbe reference

PathSeqFilterSpark Step 1: Filters low quality, low complexity, duplicate, and host reads

PathSeqPipelineSpark Combined tool that performs all steps: read filtering, microbe reference alignment, and abundance scoring

PathSeqScoreSpark Step 3: Classifies pathogen-aligned reads and generates abundance scores

--------------------------------------------------------------------------------------

Other: Miscellaneous tools, e.g. those that aid in data streaming

CreateHadoopBamSplittingIndex (BETA Tool) Create a Hadoop BAM splitting index

FifoBuffer (Picard) Provides a large, FIFO buffer that can be used to buffer input and output streams between programs.

GatherBQSRReports Gathers scattered BQSR recalibration reports into a single file

GatherTranches (BETA Tool) Gathers scattered VQSLOD tranches into a single file

IndexFeatureFile Creates an index for a feature file, e.g. VCF or BED file.

ParallelCopyGCSDirectoryIntoHDFSSpark (BETA Tool) Parallel copy a file or directory from Google Cloud Storage into the HDFS file system used by Spark

--------------------------------------------------------------------------------------

Read Data Manipulation: Tools that manipulate read data in SAM, BAM or CRAM format

AddCommentsToBam (Picard) Adds comments to the header of a BAM file.

AddOrReplaceReadGroups (Picard) Assigns all the reads in a file to a single new read-group.

ApplyBQSR Apply base quality score recalibration

ApplyBQSRSpark (BETA Tool) Apply base quality score recalibration on Spark

BQSRPipelineSpark (BETA Tool) Both steps of BQSR (BaseRecalibrator and ApplyBQSR) on Spark

BamToBfq (Picard) Converts a BAM file into a BFQ (binary fastq formatted) file

BaseRecalibrator Generates recalibration table for Base Quality Score Recalibration (BQSR)

BaseRecalibratorSpark (BETA Tool) Generate recalibration table for Base Quality Score Recalibration (BQSR) on Spark

BaseRecalibratorSparkSharded (EXPERIMENTAL Tool) BaseRecalibrator on Spark (experimental sharded implementation)

BuildBamIndex (Picard) Generates a BAM index ".bai" file.

BwaAndMarkDuplicatesPipelineSpark (BETA Tool) Takes name-sorted file and runs BWA and MarkDuplicates.

BwaSpark (BETA Tool) Align reads to a given reference using BWA on Spark

CleanSam (Picard) Cleans the provided SAM/BAM, soft-clipping beyond-end-of-reference alignments and setting MAPQ to 0 for unmapped reads

ClipReads Clip reads in a SAM/BAM/CRAM file

ConvertHeaderlessHadoopBamShardToBam (BETA Tool) Convert a headerless BAM shard into a readable BAM

DownsampleSam (Picard) Downsample a SAM or BAM file.

ExtractOriginalAlignmentRecordsByNameSpark (BETA Tool) Subsets reads by name

FastqToSam (Picard) Converts a FASTQ file to an unaligned BAM or SAM file

FilterSamReads (Picard) Subsets reads from a SAM or BAM file by applying one of several filters.

FixMateInformation (Picard) Verify mate-pair information between mates and fix if needed.

FixMisencodedBaseQualityReads Fix Illumina base quality scores in a SAM/BAM/CRAM file

GatherBamFiles (Picard) Concatenate efficiently BAM files that resulted from a scattered parallel analysis

LeftAlignIndels Left-aligns indels from reads in a SAM/BAM/CRAM file

MarkDuplicates (Picard) Identifies duplicate reads.

MarkDuplicatesGATK (EXPERIMENTAL Tool) Examines aligned records in the supplied SAM/BAM/CRAM file to locate duplicate molecules.

MarkDuplicatesSpark (BETA Tool) MarkDuplicates on Spark

MarkDuplicatesWithMateCigar (Picard) Identifies duplicate reads, accounting for mate CIGAR.

MergeBamAlignment (Picard) Merge alignment data from a SAM or BAM with data in an unmapped BAM file.

MergeSamFiles (Picard) Merges multiple SAM and/or BAM files into a single file.

PositionBasedDownsampleSam (Picard) Downsample a SAM or BAM file to retain a subset of the reads based on the reads location in each tile in the flowcell.

PrintReads Print reads in the SAM/BAM/CRAM file

PrintReadsSpark (BETA Tool) PrintReads on Spark

ReorderSam (Picard) Reorders reads in a SAM or BAM file to match ordering in a second reference file.

ReplaceSamHeader (Picard) Replaces the SAMFileHeader in a SAM or BAM file.

RevertBaseQualityScores Revert Quality Scores in a SAM/BAM/CRAM file

RevertOriginalBaseQualitiesAndAddMateCigar (Picard)Reverts the original base qualities and adds the mate cigar tag to read-group BAMs

RevertSam (Picard) Reverts SAM or BAM files to a previous state.

SamFormatConverter (Picard) Convert a BAM file to a SAM file, or a SAM to a BAM

SamToFastq (Picard) Converts a SAM or BAM file to FASTQ.

SamToFastqWithTags (Picard) Converts a SAM or BAM file to FASTQ alongside FASTQs created from tags.

SetNmAndUqTags (Picard) DEPRECATED: Use SetNmMdAndUqTags instead.

SetNmMdAndUqTags (Picard) Fixes the NM, MD, and UQ tags in a SAM file

SimpleMarkDuplicatesWithMateCigar (Picard) (EXPERIMENTAL Tool) Examines aligned records in the supplied SAM or BAM file to locate duplicate molecules.

SortSam (Picard) Sorts a SAM or BAM file

SortSamSpark (BETA Tool) SortSam on Spark (works on SAM/BAM/CRAM)

SplitNCigarReads Split Reads with N in Cigar

SplitReads Outputs reads from a SAM/BAM/CRAM by read group, sample and library name

SplitSamByLibrary (Picard) Splits a SAM or BAM file into individual files by library

SplitSamByNumberOfReads (Picard) Splits a SAM or BAM file to multiple BAMs.

UmiAwareMarkDuplicatesWithMateCigar (Picard) (EXPERIMENTAL Tool) Identifies duplicate reads using information from read positions and UMIs.

UnmarkDuplicates Clears the 0x400 duplicate SAM flag

--------------------------------------------------------------------------------------

Reference: Tools that analyze and manipulate FASTA format references

BaitDesigner (Picard) Designs oligonucleotide baits for hybrid selection reactions.

BwaMemIndexImageCreator Create a BWA-MEM index image file for use with GATK BWA tools

CreateSequenceDictionary (Picard) Creates a sequence dictionary for a reference sequence.

ExtractSequences (Picard) Subsets intervals from a reference sequence to a new FASTA file.

FindBadGenomicKmersSpark (BETA Tool) Identifies sequences that occur at high frequency in a reference

NonNFastaSize (Picard) Counts the number of non-N bases in a fasta file.

NormalizeFasta (Picard) Normalizes lines of sequence in a FASTA file to be of the same length.

ScatterIntervalsByNs (Picard) Writes an interval list created by splitting a reference at Ns.

--------------------------------------------------------------------------------------

Short Variant Discovery: Tools that perform variant calling and genotyping for short variants (SNPs, SNVs and Indels)

CombineGVCFs Merges one or more HaplotypeCaller GVCF files into a single GVCF with appropriate annotations

GenomicsDBImport Import VCFs to GenomicsDB

GenotypeGVCFs Perform joint genotyping on one or more samples pre-called with HaplotypeCaller

HaplotypeCaller Call germline SNPs and indels via local re-assembly of haplotypes

HaplotypeCallerSpark (BETA Tool) HaplotypeCaller on Spark

Mutect2 Call somatic SNVs and indels via local assembly of haplotypes

ReadsPipelineSpark (BETA Tool) Takes unaligned or aligned reads and runs BWA (if specified), MarkDuplicates, BQSR, and HaplotypeCaller to generate a VCF file of variants

--------------------------------------------------------------------------------------

Structural Variant Discovery: Tools that detect structural variants

CpxVariantReInterpreterSpark (BETA Tool) (Internal) Tries to extract simple variants from a provided GATK-SV CPX.vcf

DiscoverVariantsFromContigAlignmentsSAMSpark (BETA Tool) (Internal) Examines aligned contigs from local assemblies and calls structural variants

ExtractSVEvidenceSpark (BETA Tool) (Internal) Extracts evidence of structural variations from reads

FindBreakpointEvidenceSpark (BETA Tool) (Internal) Produces local assemblies of genomic regions that may harbor structural variants

StructuralVariationDiscoveryPipelineSpark (BETA Tool) Runs the structural variation discovery workflow on a single sample

SvDiscoverFromLocalAssemblyContigAlignmentsSpark (BETA Tool) (Internal) Examines aligned contigs from local assemblies and calls structural variants or their breakpoints

--------------------------------------------------------------------------------------

Variant Evaluation and Refinement: Tools that evaluate and refine variant calls, e.g. with annotations not offered by the engine

AnnotatePairOrientation (BETA Tool) (EXPERIMENTAL) Annotate a non-M2 VCF (using the associated tumor bam) with pair orientation fields (e.g. F1R2 ).

AnnotateVcfWithBamDepth (Internal) Annotate a vcf with a bam's read depth at each variant locus

AnnotateVcfWithExpectedAlleleFraction (Internal) Annotate a vcf with expected allele fractions in pooled sequencing

CalculateGenotypePosteriors Calculate genotype posterior probabilities given family and/or known population genotypes

CalculateMixingFractions (Internal) Calculate proportions of different samples in a pooled bam

Concordance (BETA Tool) Evaluate concordance of an input VCF against a validated truth VCF

CountFalsePositives (BETA Tool) Count PASS variants

CountVariants Counts variant records in a VCF file, regardless of filter status.

CountVariantsSpark (BETA Tool) CountVariants on Spark

FindMendelianViolations (Picard) Finds mendelian violations of all types within a VCF

Funcotator (BETA Tool) Functional Annotator

GenotypeConcordance (Picard) Calculates the concordance between genotype data of one samples in each of two VCFs - one being considered the truth (or reference) the other being the call. The concordance is broken into separate results sections for SNPs and indels. Statistics are reported in three different files.

ValidateBasicSomaticShortMutations (EXPERIMENTAL Tool) Check the variants in a VCF against a tumor-normal pair of bams representing the same samples, though not the ones from the actual calls.

ValidateVariants Validate VCF

VariantsToTable Extract fields from a VCF file to a tab-delimited table

--------------------------------------------------------------------------------------

Variant Filtering: Tools that filter variants by annotating the FILTER column

ApplyVQSR Apply a score cutoff to filter variants based on a recalibration table

CNNScoreVariants (EXPERIMENTAL Tool) Apply a Convolutional Neural Net to filter annotated variants

CNNVariantTrain (EXPERIMENTAL Tool) Train a CNN model for filtering variants

CNNVariantWriteTensors (EXPERIMENTAL Tool) Write variant tensors for training a CNN to filter variants

CreateSomaticPanelOfNormals (BETA Tool) Make a panel of normals for use with Mutect2

FilterAlignmentArtifacts (EXPERIMENTAL Tool) Filter alignment artifacts from a vcf callset.

FilterByOrientationBias (EXPERIMENTAL Tool) Filter Mutect2 somatic variant calls using orientation bias

FilterMutectCalls Filter somatic SNVs and indels called by Mutect2

FilterVariantTranches (EXPERIMENTAL Tool) Apply tranche filtering

FilterVcf (Picard) Hard filters a VCF.

VariantFiltration Filter variant calls based on INFO and/or FORMAT annotations

VariantRecalibrator Build a recalibration model to score variant quality for filtering purposes

--------------------------------------------------------------------------------------

Variant Manipulation: Tools that manipulate variant call format (VCF) data

FixVcfHeader (Picard) Replaces or fixes a VCF header.

GatherVcfs (Picard) Gathers multiple VCF files from a scatter operation into a single VCF file

GatherVcfsCloud (BETA Tool) Gathers multiple VCF files from a scatter operation into a single VCF file

LiftoverVcf (Picard) Lifts over a VCF file from one reference build to another.

MakeSitesOnlyVcf (Picard) Creates a VCF that contains all the site-level information for all records in the input VCF but no genotype information.

MakeVcfSampleNameMap (Picard) Creates a TSV from sample name to VCF/GVCF path, with one line per input.

MergeVcfs (Picard) Combines multiple variant files into a single variant file

PrintVariantsSpark (BETA Tool) Prints out variants from the input VCF.

RemoveNearbyIndels (Internal) Remove indels from the VCF file that are close to each other.

RenameSampleInVcf (Picard) Renames a sample within a VCF or BCF.

SelectVariants Select a subset of variants from a VCF file

SortVcf (Picard) Sorts one or more VCF files.

SplitVcfs (Picard) Splits SNPs and INDELs into separate files.

UpdateVCFSequenceDictionary Updates the sequence dictionary in a variant file.

UpdateVcfSequenceDictionary (Picard) Takes a VCF and a second file that contains a sequence dictionary and updates the VCF with the new sequence dictionary.

VariantAnnotator (BETA Tool) Tool for adding annotations to VCF files

VcfFormatConverter (Picard) Converts VCF to BCF or BCF to VCF.

VcfToIntervalList (Picard) Converts a VCF or BCF file to a Picard Interval List