**VarScan User's Manual**

VarScan is coded in Java, and should be executed from the command line (Terminal, in Linux/UNIX/OSX, or Command Prompt in MS Windows). For variant calling, you will need a pileup file. See the [How to Build A Pileup File](http://varscan.sourceforge.net/using-varscan.html#pileup-howto) section for details. Running VarScan with no arguments prints the usage information. Because some fields changed as of VarScan v2.2.3, we are providing updated documentations for the current release. For documentation of v2.2.2 and prior, see below.

**VarScan Documentation (v2.2.3 and later)**

USAGE: java -jar VarScan.jar [COMMAND] [OPTIONS]

COMMANDS:

**Single-sample Calling:**

[**pileup2snp**](http://varscan.sourceforge.net/using-varscan.html#v2.3_pileup2snp) [pileup file]

[**pileup2indel**](http://varscan.sourceforge.net/using-varscan.html#v2.3_pileup2indel) [pileup file]

[**pileup2cns**](http://varscan.sourceforge.net/using-varscan.html#v2.3_pileup2cns) [pileup file]

**Multi-sample Calling:**

[**mpileup2snp**](http://varscan.sourceforge.net/using-varscan.html#v2.3_mpileup2snp) [mpileup file]

[**mpileup2indel**](http://varscan.sourceforge.net/using-varscan.html#v2.3_mpileup2indel) [mpileup file]

[**mpileup2cns**](http://varscan.sourceforge.net/using-varscan.html#v2.3_mpileup2cns) [mpileup file]

**Tumor-normal Comparison:**

[**somatic**](http://varscan.sourceforge.net/using-varscan.html#v2.3_somatic) [normal pileup] [tumor pileup] or [normal-tumor mpileup]

[**copynumber**](http://varscan.sourceforge.net/using-varscan.html#v2.3_copynumber) [normal pileup] [tumor pileup] or [normal-tumor mpileup]

**Variant Filtering:**

[**filter**](http://varscan.sourceforge.net/using-varscan.html#v2.3_filter) [variants file]

[**somaticFilter**](http://varscan.sourceforge.net/using-varscan.html#v2.3_somaticFilter) [mutations file]

**Utility Functions:**

[**limit**](http://varscan.sourceforge.net/using-varscan.html#v2.3_limit) [variants file]

[**readcounts**](http://varscan.sourceforge.net/using-varscan.html#v2.3_readcounts) [pileup file]

[**compare**](http://varscan.sourceforge.net/using-varscan.html#v2.3_compare) [file1] [file2]

**pileup2snp**

This command calls SNPs from a [pileup file](http://varscan.sourceforge.net/using-varscan.html#pileup-howto) based on user-defined parameters:

USAGE: java -jar VarScan.jar pileup2snp [pileup file] OPTIONS

pileup file - The SAMtools pileup file

OPTIONS:

--min-coverage Minimum read depth at a position to make a call [8]

--min-reads2 Minimum supporting reads at a position to call variants [2]

--min-avg-qual Minimum base quality at a position to count a read [15]

--min-var-freq Minimum variant allele frequency threshold [0.01]

--p-value Default p-value threshold for calling variants [99e-02]

OUTPUT

Tab-delimited SNP calls with the following columns:

Chrom chromosome name

Position position (1-based)

Ref reference allele at this position

Cons Consensus genotype of sample in IUPAC format.

Reads1 reads supporting reference allele

Reads2 reads supporting variant allele

VarFreq frequency of variant allele by read count

Strands1 strands on which reference allele was observed

Strands2 strands on which variant allele was observed

Qual1 average base quality of reference-supporting read bases

Qual2 average base quality of variant-supporting read bases

Pvalue Significance of variant read count vs. expected baseline error

MapQual1 Average map quality of ref reads (only useful if in pileup)

MapQual2 Average map quality of var reads (only useful if in pileup)

Reads1Plus Number of reference-supporting reads on + strand

Reads1Minus Number of reference-supporting reads on - strand

Reads2Plus Number of variant-supporting reads on + strand

Reads2Minus Number of variant-supporting reads on - strand

VarAllele Most frequent non-reference allele observed

**pileup2indel**

This command calls indels from a [pileup file](http://varscan.sourceforge.net/using-varscan.html#pileup-howto) based on user-defined parameters:

USAGE: java -jar VarScan.jar pileup2indel [pileup file] OPTIONS

pileup file - The SAMtools pileup file

OPTIONS:

--min-coverage Minimum read depth at a position to make a call [8]

--min-reads2 Minimum supporting reads at a position to call variants [2]

--min-avg-qual Minimum base quality at a position to count a read [15]

--min-var-freq Minimum variant allele frequency threshold [0.01]

--p-value Default p-value threshold for calling variants [99e-02]

OUTPUT

Tab-delimited indel calls with the following columns:

Chrom chromosome name

Position position (1-based)

Ref reference allele at this position

Cons Consensus genotype of sample; \*/(var) indicates heterozygous

Reads1 reads supporting reference allele

Reads2 reads supporting variant allele

VarFreq frequency of variant allele by read count

Strands1 strands on which reference allele was observed

Strands2 strands on which variant allele was observed

Qual1 average base quality of reference-supporting read bases

Qual2 average base quality of variant-supporting read bases

Pvalue Significance of variant read count vs. expected baseline error

MapQual1 Average map quality of ref reads (only useful if in pileup)

MapQual2 Average map quality of var reads (only useful if in pileup)

Reads1Plus Number of reference-supporting reads on + strand

Reads1Minus Number of reference-supporting reads on - strand

Reads2Plus Number of variant-supporting reads on + strand

Reads2Minus Number of variant-supporting reads on - strand

VarAllele Most frequent non-reference allele observed

**pileup2cns**

This command makes consensus calls (SNP/Indel/Reference) from a [pileup file](http://varscan.sourceforge.net/using-varscan.html#pileup-howto) based on user-defined parameters:

USAGE: java -jar VarScan.jar pileup2cns [pileup file] OPTIONS

pileup file - The SAMtools pileup file

OPTIONS:

--min-coverage Minimum read depth at a position to make a call [8]

--min-reads2 Minimum supporting reads at a position to call variants [2]

--min-avg-qual Minimum base quality at a position to count a read [15]

--min-var-freq Minimum variant allele frequency threshold [0.01]

--p-value Default p-value threshold for calling variants [99e-02]

OUTPUT

Tab-delimited consensus calls with the following columns:

Chrom chromosome name

Position position (1-based)

Ref reference allele at this position

Cons Consensus genotype of sample; \*/(var) indicates heterozygous

Reads1 reads supporting reference allele

Reads2 reads supporting variant allele

VarFreq frequency of variant allele by read count

Strands1 strands on which reference allele was observed

Strands2 strands on which variant allele was observed

Qual1 average base quality of reference-supporting read bases

Qual2 average base quality of variant-supporting read bases

Pvalue Significance of variant read count vs. expected baseline error

MapQual1 Average map quality of ref reads (only useful if in pileup)

MapQual2 Average map quality of var reads (only useful if in pileup)

Reads1Plus Number of reference-supporting reads on + strand

Reads1Minus Number of reference-supporting reads on - strand

Reads2Plus Number of variant-supporting reads on + strand

Reads2Minus Number of variant-supporting reads on - strand

VarAllele Most frequent non-reference allele observed

**mpileup2snp**

This command calls SNPs from an [mpileup file](http://varscan.sourceforge.net/using-varscan.html#mpileup-howto) based on user-defined parameters:

USAGE: java -jar VarScan.jar mpileup2snp [mpileup file] OPTIONS

mpileup file - The SAMtools mpileup file

OPTIONS:

--min-coverage Minimum read depth at a position to make a call [8]

--min-reads2 Minimum supporting reads at a position to call variants [2]

--min-avg-qual Minimum base quality at a position to count a read [15]

--min-var-freq Minimum variant allele frequency threshold [0.01]

--min-freq-for-hom Minimum frequency to call homozygote [0.75]

--p-value Default p-value threshold for calling variants [99e-02]

--strand-filter Ignore variants with >90% support on one strand [1]

--output-vcf If set to 1, outputs in VCF format

--variants Report only variant (SNP/indel) positions (mpileup2cns only) [0]

OUTPUT

Tab-delimited SNP calls with the following columns:

Chrom chromosome name

Position position (1-based)

Ref reference allele at this position

Var variant allele observed

PoolCall Cross-sample call using all data (Cons:Cov:Reads1:Reads2:Freq:P-value)

Cons - consensus genotype in IUPAC format

Cov - total depth of coverage

Reads1 - number of reads supporting reference

Reads2 - number of reads supporting variant

Freq - the variant allele frequency by read count

P-value - FET p-value of observed reads vs expected non-variant

StrandFilt Information to look for strand bias using all reads (R1+:R1-:R2+:R2-:pval)

R1+ = reference supporting reads on forward strand

R1- = reference supporting reads on reverse strand

R2+ = variant supporting reads on forward strand

R2- = variant supporting reads on reverse strand

pval = FET p-value for strand distribution, R1 versus R2

SamplesRef Number of samples called reference (wildtype)

SamplesHet Number of samples called heterozygous-variant

SamplesHom Number of samples called homozygous-variant

SamplesNC Number of samples not covered / not called

SampleCalls The calls for each sample in the mpileup, space-delimited

Each sample has six values separated by colons:

Cons - consensus genotype in IUPAC format

Cov - total depth of coverage

Reads1 - number of reads supporting reference

Reads2 - number of reads supporting variant

Freq - the variant allele frequency by read count

P-value - FET p-value of observed reads vs expected non-variant

**mpileup2indel**

This command calls indels from a [mpileup file](http://varscan.sourceforge.net/using-varscan.html#mpileup-howto) based on user-defined parameters:

USAGE: java -jar VarScan.jar mpileup2indel [mpileup file] OPTIONS

mpileup file - The SAMtools mpileup file

OPTIONS:

--min-coverage Minimum read depth at a position to make a call [8]

--min-reads2 Minimum supporting reads at a position to call variants [2]

--min-avg-qual Minimum base quality at a position to count a read [15]

--min-var-freq Minimum variant allele frequency threshold [0.01]

--min-freq-for-hom Minimum frequency to call homozygote [0.75]

--p-value Default p-value threshold for calling variants [99e-02]

--strand-filter Ignore variants with >90% support on one strand [1]

--output-vcf If set to 1, outputs in VCF format

--variants Report only variant (SNP/indel) positions (mpileup2cns only) [0]

OUTPUT

Tab-delimited SNP calls with the following columns:

Chrom chromosome name

Position position (1-based)

Ref reference allele at this position

Var variant allele observed

PoolCall Cross-sample call using all data (Cons:Cov:Reads1:Reads2:Freq:P-value)

Cons - consensus genotype in IUPAC format

Cov - total depth of coverage

Reads1 - number of reads supporting reference

Reads2 - number of reads supporting variant

Freq - the variant allele frequency by read count

P-value - FET p-value of observed reads vs expected non-variant

StrandFilt Information to look for strand bias using all reads, format R1+:R1-:R2+:R2-:pval

R1+ = reference supporting reads on forward strand

R1- = reference supporting reads on reverse strand

R2+ = variant supporting reads on forward strand

R2- = variant supporting reads on reverse strand

pval = FET p-value for strand distribution, R1 versus R2

SamplesRef Number of samples called reference (wildtype)

SamplesHet Number of samples called heterozygous-variant

SamplesHom Number of samples called homozygous-variant

SamplesNC Number of samples not covered / not called

SampleCalls The calls for each sample in the mpileup, space-delimited

Each sample has six values separated by colons:

Cons - consensus genotype in IUPAC format

Cov - total depth of coverage

Reads1 - number of reads supporting reference

Reads2 - number of reads supporting variant

Freq - the variant allele frequency by read count

P-value - FET p-value of observed reads vs expected non-variant

**mpileup2cns**

This command makes consensus calls (SNP/Indel/Reference) from a [mpileup file](http://varscan.sourceforge.net/using-varscan.html#mpileup-howto) based on user-defined parameters:

USAGE: java -jar VarScan.jar mpileup2cns [mpileup file] OPTIONS

mpileup file - The SAMtools mpileup file

OPTIONS:

--min-coverage Minimum read depth at a position to make a call [8]

--min-reads2 Minimum supporting reads at a position to call variants [2]

--min-avg-qual Minimum base quality at a position to count a read [15]

--min-var-freq Minimum variant allele frequency threshold [0.01]

--min-freq-for-hom Minimum frequency to call homozygote [0.75]

--p-value Default p-value threshold for calling variants [99e-02]

--strand-filter Ignore variants with >90% support on one strand [1]

--output-vcf If set to 1, outputs in VCF format

--variants Report only variant (SNP/indel) positions (mpileup2cns only) [0]

OUTPUT

Tab-delimited SNP calls with the following columns:

Chrom chromosome name

Position position (1-based)

Ref reference allele at this position

Var variant allele observed

PoolCall Cross-sample call using all data (Cons:Cov:Reads1:Reads2:Freq:P-value)

Cons - consensus genotype in IUPAC format

Cov - total depth of coverage

Reads1 - number of reads supporting reference

Reads2 - number of reads supporting variant

Freq - the variant allele frequency by read count

P-value - FET p-value of observed reads vs expected non-variant

StrandFilt Information to look for strand bias using all reads, format R1+:R1-:R2+:R2-:pval

R1+ = reference supporting reads on forward strand

R1- = reference supporting reads on reverse strand

R2+ = variant supporting reads on forward strand

R2- = variant supporting reads on reverse strand

pval = FET p-value for strand distribution, R1 versus R2

SamplesRef Number of samples called reference (wildtype)

SamplesHet Number of samples called heterozygous-variant

SamplesHom Number of samples called homozygous-variant

SamplesNC Number of samples not covered / not called

SampleCalls The calls for each sample in the mpileup, space-delimited

Each sample has six values separated by colons:

Cons - consensus genotype in IUPAC format

Cov - total depth of coverage

Reads1 - number of reads supporting reference

Reads2 - number of reads supporting variant

Freq - the variant allele frequency by read count

P-value - FET p-value of observed reads vs expected non-variant

**somatic**

This command calls variants and identifies their somatic status (Germline/LOH/Somatic) using [pileup files](http://varscan.sourceforge.net/using-varscan.html#pileup-howto) from a matched tumor-normal pair.

USAGE: java -jar VarScan.jar somatic [normal\_pileup] [tumor\_pileup] [output] OPTIONS

normal\_pileup - The SAMtools pileup file for Normal

tumor\_pileup - The SAMtools pileup file for Tumor

output - Output base name for SNP and indel output

You can also give it a single mpileup file with normal and tumor data.

USAGE: java -jar VarScan.jar somatic [normal-tumor.mpileup] [output] --mpileup 1 OPTIONS

normal-tumor.mpileup - The SAMtools mpileup file with normal and then tumor

output - Output base name for SNP and indel output

Both formats of the command share these common options:

OPTIONS:

--output-snp - Output file for SNP calls [default: output.snp]

--output-indel - Output file for indel calls [default: output.indel]

--min-coverage - Minimum coverage in normal and tumor to call variant [8]

--min-coverage-normal - Minimum coverage in normal to call somatic [8]

--min-coverage-tumor - Minimum coverage in tumor to call somatic [6]

--min-var-freq - Minimum variant frequency to call a heterozygote [0.10]

--min-freq-for-hom Minimum frequency to call homozygote [0.75]

--normal-purity - Estimated purity (non-tumor content) of normal sample [1.00]

--tumor-purity - Estimated purity (tumor content) of tumor sample [1.00]

--p-value - P-value threshold to call a heterozygote [0.99]

--somatic-p-value - P-value threshold to call a somatic site [0.05]

--strand-filter - If set to 1, removes variants with >90% strand bias

--validation - If set to 1, outputs all compared positions even if non-variant

Note that more specific options (e.g. min-coverage-normal) will override the default or specificied value of less specific options (e.g. min-coverage).   
  
The **normal and tumor purity** values should be a value between 0 and 1. The default (1) implies that the normal is 100% pure with no contaminating tumor cells, and the tumor is 100% pure with no contaminating stromal or other non-malignant cells. You would change tumor-purity to something less than 1 if you have a low-purity tumor sample and thus expect lower variant allele frequencies for mutations. You would change normal-purity to something less than 1 only if it's possible that there will be some tumor content in your "normal" sample, e.g. adjacent normal tissue for a solid tumor, malignant blood cells in the skin punch normal for some liquid tumors, etc.   
  
There are **two p-value options**. One (p-value) is the significance threshold for the first-pass algorithm that determines, for each position, if either normal or tumor is variant at that position. The second (somatic-p-value) is more important; this is the threshold below which read count differences between tumor and normal are deemed significant enough to classify the sample as a somatic mutation or an LOH event. In the case of a shared (germline) variant, this p-value is used to determine if the combined normal and tumor evidence differ significantly enough from the null hypothesis (no variant with same coverage) to report the variant. See the [**somatic mutation calling**](http://varscan.sourceforge.net/somatic-calling.html) section for details.

OUTPUT

Two tab-delimited files (SNPs and Indels) with the following columns:

chrom chromosome name

position position (1-based from the pileup)

ref reference allele at this position

var variant allele at this position

normal\_reads1 reads supporting reference allele

normal\_reads2 reads supporting variant allele

normal\_var\_freq frequency of variant allele by read count

normal\_gt genotype call for Normal sample

tumor\_reads1 reads supporting reference allele

tumor\_reads2 reads supporting variant allele

tumor\_var\_freq frequency of variant allele by read count

tumor\_gt genotype call for Tumor sample

somatic\_status status of variant (Germline, Somatic, or LOH)

variant\_p\_value Significance of variant read count vs. baseline error rate

somatic\_p\_value Significance of tumor read count vs. normal read count

tumor\_reads1\_plus Ref-supporting reads from + strand in tumor

tumor\_reads1\_minus Ref-supporting reads from - strand in tumor

tumor\_reads2\_plus Var-supporting reads from + strand in tumor

tumor\_reads2\_minus Var-supporting reads from - strand in tumor

**copynumber**

This command calls variants and identifies their somatic status (Germline/LOH/Somatic) using [pileup files](http://varscan.sourceforge.net/using-varscan.html#pileup-howto) from a matched tumor-normal pair.

USAGE: java -jar VarScan.jar copynumber [normal\_pileup] [tumor\_pileup] [output] OPTIONS

normal\_pileup - The SAMtools pileup file for Normal

tumor\_pileup - The SAMtools pileup file for Tumor

output - Output base name for SNP and indel output

You can also give it a single mpileup file with normal and tumor data.

USAGE: java -jar VarScan.jar copynumber [normal-tumor.mpileup] [output] --mpileup 1 OPTIONS

normal-tumor.mpileup - The SAMtools mpileup file with normal and then tumor

output - Output base name for SNP and indel output

Both formats of the command share these common options:

OPTIONS:

--min-base-qual - Minimum base quality to count for coverage [20]

--min-map-qual - Minimum read mapping quality to count for coverage [20]

--min-coverage - Minimum coverage threshold for copynumber segments [20]

--min-segment-size - Minimum number of consecutive bases to report a segment [10]

--max-segment-size - Max size before a new segment is made [100]

--p-value - P-value threshold for significant copynumber change-point [0.01]

--data-ratio - The normal/tumor input data ratio for copynumber adjustment [1.0]

Note: The data ratio is intended to help you account for overall differences in the amount of sequencing coverage between normal and tumor, which might otherwise give the appearance of global copy number differences. If normal has more data than tumor, set this to something greater than 1. If tumor has more data than normal, adjust it to something below 1. A basic formula for data ratio might be something like *ratio = normal\_unique\_bp / tumor\_unique\_bp* where unique base pairs are computed as *mapped\_non\_dup\_reads \* read\_length*.

OUTPUT

chrom Chromosome name

chr\_start Region start position (1-based from the pileup)

chr\_stop Region stop position (1-based from the pileup)

num\_positions Size of the region in base pairs

normal\_depth Average normal sequence depth for the region

tumor\_depth Average tumor sequence depth for the region

log2\_ratio Log-base-2 ratio of: adjusted tumor depth over normal depth

gc\_content Estimated GC content of the region (0-100)

The raw regions reported by VarScan are delineated by drops in coverage or changes in the tumor/normal ratio, so there are many small, nearby regions with similar copy number. It is therefore recommended that raw VarScan copynumber output be processed with circular binary segmentation (CBS) or a similar algorithm, which will generate larger segments delineated by statistically significant change points. See the [**copy number calling**](http://varscan.sourceforge.net/copy-number-calling.html) section for details.

**filter**

This command filters variants in a file by coverage, supporting reads, variant frequency, or average base quality. It is for use with output from pileup2snp or pileup2indel.

USAGE: java -jar VarScan.jar filter [variants file] OPTIONS

variants file - A file of SNP or indel calls from VarScan pileup2snp or pileup2indel

OPTIONS:

--min-coverage Minimum read depth at a position to make a call [10]

--min-reads2 Minimum supporting reads at a position to call variants [2]

--min-strands2 Minimum # of strands on which variant observed (1 or 2) [1]

--min-avg-qual Minimum average base quality for variant-supporting reads [20]

--min-var-freq Minimum variant allele frequency threshold [0.20]

--p-value Default p-value threshold for calling variants [1e-01]

--indel-file File of indels for filtering nearby SNPs, from pileup2indel command

--output-file File to contain variants passing filters

**somaticFilter**

This command filters somatic mutation calls to remove clusters of false positives and SNV calls near indels. Note: this is a basic filter. More advanced filtering strategies consider mapping quality, read mismatches, soft-trimming, and other factors when deciding whether or not to filter a variant. See the [VarScan 2 publication](http://www.ncbi.nlm.nih.gov/pubmed/19542151) (Koboldt et al, Genome Research, Feb 2012) for details.

USAGE: java -jar VarScan.jar somaticFilter [mutations file] OPTIONS

mutations file - A file of SNVs from VarScan somatic

OPTIONS:

--min-coverage Minimum read depth [10]

--min-reads2 Minimum supporting reads for a variant [2]

--min-strands2 Minimum # of strands on which variant observed (1 or 2) [1]

--min-avg-qual Minimum average base quality for variant-supporting reads [20]

--min-var-freq Minimum variant allele frequency threshold [0.20]

--p-value Default p-value threshold for calling variants [1e-01]

--indel-file File of indels for filtering nearby SNPs

--output-file Optional output file for filtered variants

**limit**

This command limits variants in a file to a set of positions or regions

USAGE: java -jar VarScan.jar limit [infile] OPTIONS

infile - A file of chromosome-positions, tab-delimited

OPTIONS

--positions-file - a file of chromosome-positions, tab delimited

--regions-file - a file of chromosome-start-stops, tab delimited

--output-file - Output file for the matching variants

**readcounts**

This command reports the read counts for each base at positions in a pileup file

USAGE: java -jar VarScan.jar readcounts [pileup file] OPTIONS

pileup file - The SAMtools pileup file

OPTIONS:

--variants-file A list of variants at which to report readcounts

--output-file Output file to contain the readcounts

--min-coverage Minimum read depth at a position to make a call [8]

--min-base-qual Minimum base quality at a position to count a read [30]

**compare**

This command performs set-comparison operations on two files of variants.

USAGE: java -jar VarScan.jar compare [file1] [file2] [type] [output] OPTIONS

file1 - A file of chromosome-positions, tab-delimited

file2 - A file of chromosome-positions, tab-delimited

type - Type of comparison [intersect|merge|unique1|unique2]

output - Output file for the comparison result

For detailed usage information, see the [VarScan JavaDoc](http://varscan.sourceforge.net/doc/index.html).

**VarScan Documentation (v2.2.2 and before)**

USAGE: java -jar VarScan.jar [COMMAND] [OPTIONS]

COMMANDS

[**pileup2snp**](http://varscan.sourceforge.net/using-varscan.html#v2.2_pileup2snp) [pileup file]

[**pileup2indel**](http://varscan.sourceforge.net/using-varscan.html#v2.2_pileup2indel) [pileup file]

[**pileup2cns**](http://varscan.sourceforge.net/using-varscan.html#v2.2_pileup2cns) [pileup file]

[**somatic**](http://varscan.sourceforge.net/using-varscan.html#v2.2_somatic) [normal pileup] [tumor pileup]

[**filter**](http://varscan.sourceforge.net/using-varscan.html#v2.2_filter) [variants file]

[**somaticFilter**](http://varscan.sourceforge.net/using-varscan.html#v2.2_somaticFilter) [mutations file]

[**limit**](http://varscan.sourceforge.net/using-varscan.html#v2.2_limit) [variants file]

[**readcounts**](http://varscan.sourceforge.net/using-varscan.html#v2.2_readcounts) [pileup file]

[**compare**](http://varscan.sourceforge.net/using-varscan.html#v2.2_compare) [file1] [file2]

**pileup2snp**

This command calls SNPs from a [pileup file](http://varscan.sourceforge.net/using-varscan.html#pileup-howto) based on user-defined parameters:

USAGE: java -jar VarScan.jar pileup2snp [pileup file] OPTIONS

pileup file - The SAMtools pileup file

OPTIONS:

--min-coverage Minimum read depth at a position to make a call [10]

--min-reads2 Minimum supporting reads at a position to call variants [2]

--min-avg-qual Minimum base quality at a position to count a read [15]

--min-var-freq Minimum variant allele frequency threshold [0.01]

--p-value Default p-value threshold for calling variants [99e-02]

OUTPUT

Tab-delimited SNP calls with the following columns:

Chrom chromosome name

Position position (1-based)

Ref reference allele at this position

Var variant allele at this position

Reads1 reads supporting reference allele

Reads2 reads supporting variant allele

VarFreq frequency of variant allele by read count

Strands1 strands on which reference allele was observed

Strands2 strands on which variant allele was observed

Qual1 average base quality of reference-supporting read bases

Qual2 average base quality of variant-supporting read bases

Pvalue Significance of variant read count vs. expected baseline error

**pileup2indel**

This command calls indels from a [pileup file](http://varscan.sourceforge.net/using-varscan.html#pileup-howto) based on user-defined parameters:

USAGE: java -jar VarScan.jar pileup2indel [pileup file] OPTIONS

pileup file - The SAMtools pileup file

OPTIONS:

--min-coverage Minimum read depth at a position to make a call [8]

--min-reads2 Minimum supporting reads at a position to call variants [2]

--min-avg-qual Minimum base quality at a position to count a read [15]

--min-var-freq Minimum variant allele frequency threshold [0.01]

--p-value Default p-value threshold for calling variants [99e-02]

OUTPUT

Tab-delimited indel calls with the following columns:

Chrom chromosome name

Position position (1-based)

Ref reference allele at this position

Var variant allele at this position

Reads1 reads supporting reference allele

Reads2 reads supporting variant allele

VarFreq frequency of variant allele by read count

Strands1 strands on which reference allele was observed

Strands2 strands on which variant allele was observed

Qual1 average base quality of reference-supporting read bases

Qual2 average base quality of variant-supporting read bases

Pvalue Significance of variant read count vs. expected baseline error

**pileup2cns**

This command makes consensus calls (SNP/Indel/Reference) from a [pileup file](http://varscan.sourceforge.net/using-varscan.html#pileup-howto) based on user-defined parameters:

USAGE: java -jar VarScan.jar pileup2cns [pileup file] OPTIONS

pileup file - The SAMtools pileup file

OPTIONS:

--min-coverage Minimum read depth at a position to make a call [8]

--min-reads2 Minimum supporting reads at a position to call variants [2]

--min-avg-qual Minimum base quality at a position to count a read [15]

--min-var-freq Minimum variant allele frequency threshold [0.01]

--p-value Default p-value threshold for calling variants [99e-02]

OUTPUT

Tab-delimited consensus calls with the following columns:

Chrom chromosome name

Position position (1-based)

Ref reference allele at this position

Var consensus call (reference, IUPAC SNP code, or indel)

Reads1 reads supporting reference allele

Reads2 reads supporting variant allele

VarFreq frequency of variant allele by read count

Strands1 strands on which reference allele was observed

Strands2 strands on which variant allele was observed

Qual1 average base quality of reference-supporting read bases

Qual2 average base quality of variant-supporting read bases

Pvalue Significance of variant read count vs. expected baseline error

**somatic**

This command calls variants and identifies their somatic status (Germline/LOH/Somatic) using [pileup files](http://varscan.sourceforge.net/using-varscan.html#pileup-howto) from a matched tumor-normal pair.

USAGE: java -jar VarScan.jar somatic [normal\_pileup] [tumor\_pileup] [output] OPTIONS

normal\_pileup - The SAMtools pileup file for Normal

tumor\_pileup - The SAMtools pileup file for Tumor

output - Output base name for SNP and indel output

OPTIONS:

--output-snp Output file for SNP calls [output.snp]

--output-indel Output file for indel calls [output.indel]

--min-coverage Minimum coverage in normal and tumor to call variant [10]

--min-coverage-normal Minimum coverage in normal to call somatic [10]

--min-coverage-tumor Minimum coverage in tumor to call somatic [5]

--min\_var\_freq Minimum variant frequency to call a heterozygote [0.20]

--p-value P-value threshold to call a heterozygote [1.0e-01]

--somatic-p-value P-value threshold to call a somatic site [1.0e-04]

OUTPUT

Two tab-delimited files (SNPs and Indels) with the following columns:

Chrom chromosome name

Position position (1-based)

Ref reference allele at this position

Var variant allele at this position

Normal\_Reads1 reads supporting reference allele

Normal\_Reads2 reads supporting variant allele

Normal\_VarFreq frequency of variant allele by read count

Normal\_Gt genotype call for Normal sample

Tumor\_Reads1 reads supporting reference allele

Tumor\_Reads2 reads supporting variant allele

Tumor\_VarFreq frequency of variant allele by read count

Tumor\_Gt genotype call for Tumor sample

Somatic\_Status status of variant (Germline, Somatic, or LOH)

Pvalue Significance of variant read count vs. expected baseline error

Somatic\_Pvalue Significance of tumor read count vs. normal read count

**filter**

This command filters variants in a file by coverage, supporting reads, variant frequency, or average base quality

USAGE: java -jar VarScan.jar filter [variants file] OPTIONS

variants file - A file of SNP or indel calls from VarScan

OPTIONS:

--min-coverage Minimum read depth at a position to make a call [8]

--min-reads2 Minimum supporting reads at a position to call variants [2]

--min-avg-qual Minimum base quality at a position to count a read [15]

--min-var-freq Minimum variant allele frequency threshold [0.01]

--p-value Default p-value threshold for calling variants [99e-02]

**somaticFilter**

This command filters somatic mutation calls to remove clusters of false positives and SNV calls near indels.

USAGE: java -jar VarScan.jar somaticFilter [mutations file] OPTIONS

mutations file - A file of SNVs from VarScan somatic

OPTIONS:

--min-coverage Minimum read depth [10]

--min-reads2 Minimum supporting reads for a variant [2]

--min-strands2 Minimum # of strands on which variant observed (1 or 2) [1]

--min-avg-qual Minimum average base quality for variant-supporting reads [20]

--min-var-freq Minimum variant allele frequency threshold [0.20]

--p-value Default p-value threshold for calling variants [1e-01]

--indel-file File of indels for filtering nearby SNPs

--output-file Optional output file for filtered variants

**limit**

This command limits variants in a file to a set of positions or regions

USAGE: java -jar VarScan.jar limit [infile] OPTIONS

infile - A file of chromosome-positions, tab-delimited

OPTIONS

--positions-file - a file of chromosome-positions, tab delimited

--regions-file - a file of chromosome-start-stops, tab delimited

--output-file - Output file for the matching variants

**readcounts**

This command reports the read counts for each base at positions in a pileup file

USAGE: java -jar VarScan.jar readcounts [pileup file] OPTIONS

pileup file - The SAMtools pileup file

OPTIONS:

--variants-file A list of variants at which to report readcounts

--output-file Output file to contain the readcounts

--min-coverage Minimum read depth at a position to make a call [8]

--min-base-qual Minimum base quality at a position to count a read [30]

**compare**

This command performs set-comparison operations on two files of variants.

USAGE: java -jar VarScan.jar compare [file1] [file2] [type] [output] OPTIONS

file1 - A file of chromosome-positions, tab-delimited

file2 - A file of chromosome-positions, tab-delimited

type - Type of comparison [intersect|merge|unique1|unique2]

output - Output file for the comparison result

For detailed usage information, see the [VarScan JavaDoc](http://varscan.sourceforge.net/doc/index.html).

**How to Build a SAMtools (m)pileup File**

The variant calling features of VarScan for single samples (pileup2snp, pileup2indel, pileup2cns) and multiple samples (mpileup2snp, mpileup2indel, mpileup2cns, and somatic) expect input in SAMtools [pileup](http://samtools.sourceforge.net/mpileup.shtml) or [mpileup](http://samtools.sourceforge.net/mpileup.shtml) format. **In current versions of SAMtools, the "pileup" command has now been replaced with the "mpileup" command**. For a single sample, these operate in a very similar fashion, except that mpileup applies BAQ adjustments by default, and the output is identical. When you give it multiple BAM files, however, SAMtools mpileup generates a multi-sample pileup format that must be processed with the mpileup2\* commands in VarScan. To build a mpileup file, you will need:

* One or more **BAM** files ("myData.bam") that have been sorted using the *sort* command of SAMtools.
* The **reference sequence** ("reference.fasta") to which reads were aligned, in FASTA format.
* The [**SAMtools**](http://samtools.sourceforge.net/) software package.

Generate a mpileup file with the following command:

samtools mpileup -f [reference sequence] [BAM file(s)] >myData.mpileup

Note, to save disk space and file I/O, you can redirect mpileup output directly to VarScan with a "pipe" command. For example:

One sample:

samtools mpileup -f reference.fasta myData.bam | java -jar VarScan.v2.2.jar pileup2snp

Multiple samples:

samtools mpileup -f reference.fasta sample1.bam sample2.bam | java -jar VarScan.v2.2.jar pileup2snp