Package: methylVCF (via r-universe)

November 22, 2026

Title Query DNA Methylation in VCF and gVCF Files

Version 0.99.0

Description Queries DNA methylation by genomic region in Illumina DRAGEN 5-Base VCF and gVCF files and in Rastair VCF files. Returns genomic ranges with the reported methylation fraction, informative read depth, and allele or haplotype assignment when supported by the input. Preserves source record coordinates and distinguishes zero informative depth from an observed zero methylation fraction. Aligns reference methylation measurements across samples in a ranged experiment.

License MIT + file LICENSE

URL https://github.com/xec-cm/methylVCF, https://xec-cm.github.io/methylVCF/

BugReports https://github.com/xec-cm/methylVCF/issues

Encoding UTF-8

Depends R (>= 4.5)

Imports cli (>= 3.0.0), GenomicRanges, IRanges, Rsamtools, S4Vectors, Seqinfo, stats, SummarizedExperiment, utils

Suggests BiocStyle, dplyr, knitr, plyranges, rmarkdown, testthat (>= 3.2.0), tibble, tidySummarizedExperiment

Config/testthat/edition 3

Roxygen list(roclets = c(``namespace'', ``rd'', ``roxytest::testthat_roclet''))

VignetteBuilder knitr

biocViews DataImport, DNAMethylation, Epigenetics, MethylSeq, Sequencing, Coverage

Config/roxygen2/version 8.1.1

Config/pak/sysreqs libbz2-dev liblzma-dev xz-utils zlib1g-dev

Repository https://biocstaging.r-universe.dev

Date/Publication 2026-11-21 20:03:09 UTC

RemoteUrl https://github.com/BiocStaging/methylVCF

RemoteRef HEAD

RemoteSha 873f6bde7f164e96f17bc9dd4ad5e69bbdace409

Contents

asMethylExperiment 2
inspectMethylVcf 3
readMethylVcf 4

Index **7**

`asMethylExperiment` *Align reference methylation calls across samples*

Description

Project regional calls onto a shared set of reference sites. DRAGEN inputs contribute pooled reference-allele measurements. Rastair inputs contribute reference CpG context measurements. All inputs must use the same format and reference genome.

Usage

```
asMethylExperiment(calls, filters = c("PASS", "."))
```

Arguments

`calls`	A `GRanges` from `readMethylVcf()` containing one or more VCF samples, or a named list of these objects with one sample per element. List names become column names. Use the reader's `samples` argument to select a single sample from a multisample file for a list element.
`filters`	Exact VCF FILTER values to retain. By default, retain `"PASS"` and `"."` (filters not applied). Use `NULL` to include all records.

Details

The `fraction` assay contains fractions backed by positive informative depth. The `reported_fraction` assay preserves the raw VCF values. `informative_depth`, `modified_depth` and `coverage_status` preserve the original values independently. All assays use `NA` where there is no call for a site and sample. A reported zero-depth call keeps its depth of zero and its coverage status.

Strand-specific calls and collapsed dyads occupy separate rows. This projection excludes alternate, de novo, allele-specific and haplotype measurements. Use the original ranges to analyze those measurements. Multiple calls that pass the `filters` selection at one site in one sample signal an error with class `methylVCF_duplicate_site`.

Conflicting assemblies or sequence lengths signal an error. Different reference file locations are allowed when the inputs declare a common assembly. Other conflicting `##reference` declarations signal an error. Confirm a common assembly when header information is incomplete. The assays use dense matrices intended for regional data.

Value

A SummarizedExperiment::RangedSummarizedExperiment with one row per chromosome, 1-based anchor, strand, context, and modification across the union of selected sites. Each selected sample has a column, including samples without calls. `colData()` stores the original VCF sample ID and an `import_index` into `S4Vectors::metadata(result)$imports`, which retains the original imports' file metadata and diagnostics. The experiment metadata also record the chosen `filters`.

See Also

`readMethylVcf()` for importing the individual measurements and `vignette("methylVCF")` for the complete workflow.

Examples

```
file <- system.file(
    "extdata", "dragen-multisample.g.vcf", package = "methylVCF"
)
region <- GenomicRanges::GRanges("chr21", IRanges::IRanges(13204474L))
calls <- readMethylVcf(file, region)
experiment <- asMethylExperiment(calls)
SummarizedExperiment::assay(experiment, "fraction")
SummarizedExperiment::assay(experiment, "reported_fraction")
```

---

inspectMethylVcf *Inspect a methylation VCF before importing it*

Description

Read the header of a local DRAGEN or Rastair VCF and optionally check whether its index opens. The inspection stops at the column header and does not decode methylation records.

Usage

```
inspectMethylVcf(file, format = NULL, check_index = TRUE)
```

Arguments

`file`	Path to a local VCF or gVCF, optionally compressed.
`format`	The expected caller, `"dragen"` or `"rastair"`. With `NULL`, identify the caller from the header when possible.
`check_index`	Logical. Try to open the index that `readMethylVcf()` would select. Set to `FALSE` to check only for its presence.

Details

The selected CSI index takes precedence over TBI, including when the CSI index is unreadable. A readable index can be opened and its contigs listed. Its relationship to the VCF and the VCF records themselves remain to be checked during a regional import.

A missing methylation declaration is reported in `checks`. A path that cannot be opened or a structurally invalid VCF header raises an error. The checks do not infer missing reference builds or require a particular declared methylation-vector cardinality.

Value

A list of class `methylVCF_inspection` with a console summary and the following components:

- `file`, `caller`, `source`, `vcf_version`, `caller_version` and `reference` describe the input and its declared provenance. Unavailable values are `NA`. `format` is the caller used for the header checks.
- `samples` contains VCF sample IDs in file order. `seqinfo` holds the contigs, lengths and any assembly tags declared per contig. The file-level reference declaration is returned as `reference`.
- `fields` is a data frame of INFO and FORMAT declarations with `section`, `id`, `number` and `type` columns.
- `checks` is a data frame of header findings with `check`, `status` (`"pass"`, `"warning"` or `"error"`) and `message` columns.
- `index` is a list with the chosen `path`, `type` (`"TBI"` or `"CSI"`), `status` (`"absent"`, `"detected"`, `"readable"` or `"unreadable"`), indexed `contigs`, and a `message` if opening fails.

See Also

`readMethylVcf()` to query and interpret the records.

Examples

```
file <- system.file("extdata", "dragen-multisample.g.vcf", package = "methylVCF")
info <- inspectMethylVcf(file)
info
info$samples
```

---

readMethylVcf *Query reported 5mC in a genomic region*

Description

Interpret 5mC fields from DRAGEN and Rastair and retain evidence supplied by the file, including explicit missing slots and zero fractions. The query filters observations by their reported anchors after decoding.

Usage

```
readMethylVcf(
  file,
  which,
  format = c("dragen", "rastair"),
  samples = NULL,
  on_problem = c("error", "report")
)
```

Arguments

file	Path to a local text VCF or gVCF, optionally compressed with gzip or BGZF.
which	A GRanges with 1-based inclusive regions to query. The reader combines overlapping regions and ignores strand in the query.
format	The producer's format, "dragen" (default) or "rastair". Use the format reported by inspectMethylVcf() when identified.
samples	Character vector of VCF sample names to return, or NULL to retain all samples.
on_problem	"error" stops at the first unsupported or invalid record, sample, or measurement. "report" retains readable observations and records omissions in the result metadata. Header, file, and index errors always stop the import.

Details

DRAGEN support includes reference blocks with INFO/END, diploid SNVs including multiallelic sites, and measurements on mappable alleles of complex records. When REF and ALT differ in length, projection is limited to their shared leading anchor. Rastair reference and biallelic single-base contexts use their producer-specific order. When both CPG and CPGnovo are set, values must be separable by context. Rastair 2.2.0 can emit a single slot for such records. The reader reports that representation as unsupported. A Rastair slot with more modified reads (ADM5mC) than informative reads (DPM5mC) is invalid when both depths are available.

scope is "pooled", "allele", "haplotype", or "context". A DRAGEN phased homozygote with one measurement per site has pooled scope and a missing haplotype. A Rastair allele assignment identifies a local sequence allele to which the CpG context can be attributed. Its haplotype is missing. If multiple local alleles could anchor the context, the reader retains the measurement with a missing allele index and unspecified strand. Unreported measurements remain missing. The file-level metadata include the chosen format, selected sample IDs (even for samples without calls), and available header provenance.

coverage_status is "observed" for a reported fraction with positive informative depth. "no_informative_depth" takes precedence when the depth is zero. "missing_fraction" indicates a missing fraction and "unknown_depth" indicates a reported fraction with missing depth. fraction retains the producer's value.

The reader uses Rsamtools for BGZF files with a companion TBI or CSI index, preferring CSI when both are present. Otherwise it scans the input sequentially in chunks. Indexed queries retrieve the requested intervals and decode a source record only in its first overlapping interval, keeping all measurements within the query. Anchors include the start of a collapsed CpG dyad. Retained rows preserve source order within each record.

Value

A GRanges with one row per reported measurement slot and a one-base range at its reported anchor. Metadata columns include the sample, context, scope, fraction, informative and modified read depths, genotype, optional haplotype and phase-set identifiers, allele index (0 for REF and 1 for the first ALT), and the source record coordinates. allele_offset is the zero-based offset within the allele sequence or reference block. In a gVCF block, allele gives the sequence at the record's anchor only. methylation_mask retains the original DRAGEN INFO/M5mC string in REF/ALT order when present. The Seqinfo retains all contigs, lengths and assembly tags declared in the VCF header. Queried contigs absent from the header have unknown lengths and assemblies.

Import diagnostics

S4Vectors::metadata(result)$diagnostics contains one row per import issue, with chrom, pos, sample when known, type ("invalid" or "unsupported"), allele_index, allele_offset, and value_index when available, and reason. In "error" mode, the first such issue signals a methylVCF_import_error condition with those fields. In "report" mode, a single methylVCF_import_warning condition carries the total issue count in n_issues.

See Also

inspectMethylVcf() for checking the header and index, asMethylExperiment() for comparing reference sites across samples, and vignette("methylVCF") for the complete workflow.

Examples

```
file <- system.file("extdata", "dragen-example.g.vcf", package = "methylVCF")
region <- GenomicRanges::GRanges("chr21", IRanges::IRanges(13204474L))
result <- readMethylVcf(file, region, format = "dragen")
result
```

Index

asMethylExperiment, 2
asMethylExperiment(), *6*

inspectMethylVcf, 3
inspectMethylVcf(), *5*, *6*

readMethylVcf, 4
readMethylVcf(), *2–4*

SummarizedExperiment::RangedSummarizedExperiment, *3*