

# Package: gsalib (via r-universe)

October 10, 2024

**Type** Package

**Title** Utility Functions for 'GATK'

**Version** 2.2.1

**Date** 2022-10-07

**Author** Kiran Garimella

**Maintainer** Louis Bergelson <louisb@broadinstitute.org>

**Description** Provides utility functions used by the Genome Analysis Toolkit ('GATK') to load tables and plot data. The 'GATK' is a toolkit for variant discovery in high-throughput sequencing data.

**URL** <https://gatk.broadinstitute.org/hc/en-us>,  
<https://github.com/broadinstitute/gatk>,  
<https://github.com/broadinstitute/gsalib/>

**BugReports** <https://github.com/broadinstitute/gsalib/issues>

**License** MIT + file LICENSE

**LazyLoad** yes

**NeedsCompilation** no

**Date/Publication** 2013-07-03 17:30:49

**Repository** <https://broadinstitute.r-universe.dev>

**RemoteUrl** <https://github.com/broadinstitute/gsalib>

**RemoteRef** HEAD

**RemoteSha** f1822a3f2a3b4afed31edffff35dd5dac2136731

## Contents

gsalib-package	2
gsa.read.gatkreport	2
gsa.read.gatkreportv0	3
gsa.read.gatkreportv1	4
gsa.reshape.concordance.table	5

**Index**

7

---

gsalib-package	<i>Utility functions for GATK</i>
----------------	-----------------------------------

---

**Description**

Utility functions for analysis of genome sequence data with the GATK

**Details**

This package is primarily meant to be used programmatically by GATK tools. However the `gsa.read.gatkreport()` function can be used to easily read in data from a GATKReport. A GATKReport is a multi-table document generated by GATK tools.

**Author(s)**

Kiran Garimella

Maintainer: Louis Bergelson <louisb@broadinstitute.org>

**References**

<https://gatk.broadinstitute.org/hc/en-us/articles/360035532172-GATKReport-and-gsalib>

**Examples**

```
test_file = system.file("inst", "extdata", "test_gatkreport.table", package = "gsalib");
report = gsa.read.gatkreport(test_file);
```

---

gsa.read.gatkreport	<i>Function to read in a GATKReport</i>
---------------------	---

---

**Description**

This function reads in data from a GATKReport. A GATKReport is a document containing multiple tables produced by the GATK. Each table is loaded as a separate `data.frame` object in a list.

**Usage**

```
gsa.read.gatkreport(filename)
```

**Arguments**

`filename` The path to the GATKReport file.

## Details

The GATKReport format replaces the multi-file output format used previously by many GATK tools and provides a single, consolidated file format. This format accommodates multiple tables and is still R-loadable through this function.

## Value

Returns a LIST object, where each key is the TableName and the value is the data.frame object with the contents of the table. If multiple tables with the same name exist, each one after the first will be given names of TableName.v1, TableName.v2, ..., TableName.vN.

## Note

This function accepts different versions of the GATKReport format by making internal calls to gsa.read.gatkreportv0() or gsa.read.gatkreportv1() as appropriate.

## Author(s)

Kiran Garimella

## References

<https://gatk.broadinstitute.org/hc/en-us/articles/360035532172-GATKReport-and-gsalib>

## Examples

```
test_file = system.file("extdata", "test_gatkreport.table", package = "gsalib");
report = gsa.read.gatkreport(test_file);
```

---

`gsa.read.gatkreportv0` *Function to read in an old-style GATKReport*

---

## Description

This function reads in data from a version 0.x GATKReport. It should not be called directly; instead, use gsa.read.gatkreport()

## Usage

```
gsa.read.gatkreportv0(lines)
```

## Arguments

`lines`            The lines read in from the input file.

**Value**

Returns a LIST object, where each key is the TableName and the value is the data.frame object with the contents of the table. If multiple tables with the same name exist, each one after the first will be given names of TableName.v1, TableName.v2, ..., TableName.vN.

**Author(s)**

Kiran Garimella

**References**

<https://gatk.broadinstitute.org/hc/en-us/articles/360035532172-GATKReport-and-gsalib>

---

gsa.read.gatkreportv1 *Function to read in a new-style GATKReport*

---

**Description**

This function reads in data from a version 1.x GATKReport. It should not be called directly; instead, use gsa.read.gatkreport()

**Usage**

```
gsa.read.gatkreportv1(lines)
```

**Arguments**

lines            The lines read in from the input file.

**Value**

Returns a LIST object, where each key is the TableName and the value is the data.frame object with the contents of the table. If multiple tables with the same name exist, each one after the first will be given names of TableName.v1, TableName.v2, ..., TableName.vN.

**Author(s)**

Kiran Garimella

**References**

<https://gatk.broadinstitute.org/hc/en-us/articles/360035532172-GATKReport-and-gsalib>

---

`gsa.reshape.concordance.table`*Reshape a Concordance Table*

---

### Description

Given a GATKReport generated by GenotypeConcordance (as output by `gsa.read.gatkreport`), this function reshapes the concordance for a specified sample into a matrix with the EvalGenotypes in rows and the CompGenotypes in columns (see the documentation for GenotypeConcordance for the definition of Eval and Comp)

### Usage

```
gsa.reshape.concordance.table(  
  report,  
  table.name="GenotypeConcordance_Counts",  
  sample.name="ALL")
```

### Arguments

<code>report</code>	A GATKReport as output by <code>gsa.read.gatkreport</code> . If <code>table.name</code> is NULL, <code>report</code> is assumed to be the vector of concordance values to reshape.
<code>table.name</code>	The table name in the GATKReport to reshape. Defaults to "GenotypeConcordance_Counts", but could also be one of the proportion tables ("GenotypeConcordance_EvalProportions", "GenotypeConcordance_CompProportions"). This value can also be NULL, in which case <code>report</code> is reshaped directly.
<code>sample.name</code>	The sample name within <code>table.name</code> to use.

### Value

Returns a two-dimensional matrix with Eval genotypes in the rows and Comp genotypes in the columns. The genotypes themselves (HOM\_REF, NO\_CALL, etc) are specified in the row/col names of the matrix.

### Author(s)

Phillip Dexheimer

### See Also

[gsa.read.gatkreport](#)

**Examples**

```
test_file = system.file("extdata", "test_genconcord.table", package = "gsalib")
report = gsa.read.gatkreport(test_file)
gsa.reshape.concordance.table(report)
```

```
## Output looks like:
```

```
##                               CompGenotypes
##EvalGenotypes NO_CALL HOM_REF HET HOM_VAR UNAVAILABLE MIXED
## NO_CALL      0         0       0         0         0         0
## HOM_REF      0         0       0         0         0         0
## HET          0         0     13463     90       3901       0
## HOM_VAR      0         0     2935     18144   4448       0
## UNAVAILABLE  0         0    2053693 1326112 11290       0
## MIXED        0         0         0         0         0         0
```

# Index

## \* **manip**

- gsa.read.gatkreport, [2](#)
- gsa.read.gatkreportv0, [3](#)
- gsa.read.gatkreportv1, [4](#)
- gsa.reshape.concordance.table, [5](#)

## \* **package**

- gsalib-package, [2](#)

- gsa.read.gatkreport, [2](#), [5](#)
- gsa.read.gatkreportv0, [3](#)
- gsa.read.gatkreportv1, [4](#)
- gsa.reshape.concordance.table, [5](#)
- gsalib (gsalib-package), [2](#)
- gsalib-package, [2](#)