Package ‘savR’

March 23, 2017

Type  Package
Title  Parse and analyze Illumina SAV files
Version  1.12.0
Date  2015-07-28
Author  R. Brent Calder
Maintainer  R. Brent Calder <brent.calder@einstein.yu.edu>
Description  Parse Illumina Sequence Analysis Viewer (SAV) files, access data, and generate QC plots.
License  AGPL-3
URL  https://github.com/bcalder/savR
BugReports  https://github.com/bcalder/savR/issues
Depends  ggplot2
Imports  methods, reshape2, scales, gridExtra, XML
Suggests  Cairo, testthat
biocViews  Sequencing
NeedsCompilation  no

R topics documented:

savR-package ......................................................... 2
buildReports ..................................................... 3
clusterQualityGtN ................................................ 3
clusters ............................................................. 4
correctedIntensities ............................................ 5
cycles ............................................................... 6
directions ........................................................... 6
errorMetrics ........................................................ 7
extractionMetrics .................................................. 7
flowcellLayout ...................................................... 8
illuminaFlowCellLayout-class .................................... 9
illuminaRead-class ............................................... 9
location ............................................................ 9
pfBoxplot .......................................................... 10
pfClusters ......................................................... 10
**Description**

Parse Illumina Sequence Analysis Viewer files

**Details**

<table>
<thead>
<tr>
<th>Package</th>
<th>savR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>Package</td>
</tr>
<tr>
<td>Version</td>
<td>1.7.5</td>
</tr>
<tr>
<td>Date</td>
<td>2015-07-28</td>
</tr>
<tr>
<td>License</td>
<td>AGPL-3</td>
</tr>
<tr>
<td>LazyLoad</td>
<td>yes</td>
</tr>
</tbody>
</table>

Parse Illumina Sequence Analysis Viewer (SAV) files, access data, and generate QC plots.

**Author(s)**

R. Brent Calder <brent.calder@einstein.yu.edu>

**References**

For information about Illumina SAV, please refer to
sequencinganalysisviewer_userguide_15020619c.pdf

For other implementations (and inspiration) please see
http://search.cpan.org/dist/Bio-IlluminaSAV/Bio/IlluminaSAV.pm
https://bitbucket.org/invitae/illuminate
buildReports  Generate Illumina reports folder

Description

Generate a folder of images that approximates the format of the folder that was superceded by InterOp. Requires the Cairo package.

Usage

buildReports(project, destination)

## S4 method for signature 'savProject, character'
buildReports(project,
    destination = "./savR-reports")

## S4 method for signature 'savProject, missing'
buildReports(project)

Arguments

project  SAV project
destination  path to save reports folder

Examples

## Not run:
example(savR)
buildReports(fc, "reports")

## End(Not run)

clusterQualityGtN  Get the proportion of clusters over a specified quality threshold

Description

Return the ratio of clusters with a quality score less than or equal to a specified value (n) for the requested lanes and cycles.

Usage

clusterQualityGtN(project, lane, cycle, n)

## S4 method for signature 'savProject, integer, integer, integer'
clusterQualityGtN(project, lane,
    cycle, n = 30L)
clusters

Arguments

- **project**: SAV project
- **lane**: lane(s) number
- **cycle**: cycle(s) number
- **n**: quality threshold

Examples

```r
## Not run:
example(savR)
clusterQualityGtN(fc, 1L, 25L, 30L)
## End(Not run)
```

---

clusters

*Get number of clusters per lane*

Description

Sum the total number of clusters for all tiles in the lane.

Usage

```r
clusters(project, lane)
```

- **S4 method for signature 'savProject, integer'**
  ```r
  clusters(project, lane = 1L)
  ```

Arguments

- **project**: SAV project
- **lane**: lane(s) number

Examples

```r
## Not run:
example(savR)
clusters(fc, 1L)
## End(Not run)
```
correctedIntensities  

Get Corrected Intensity data

Description

Returns a data frame of corrected intensity data.

Usage

correctedIntensities(project)

## S4 method for signature 'savProject'
correctedIntensities(project)

Arguments

project  SAV project

Details

lane: Lane number

tile: Tile ID

cycle: Cycle number

avg_intensity: Average intensity

avg_cor_[ACGT]: Average corrected intensity of channel A, C, G, or T

avg_cor_called_[ACGT]: Average corrected intensity for called clusters in channel A, C, G, or T

num_{none|[ACGT]}: Number of called bases for no-call, A, C, G, or T

sig_noise: Signal to noise ratio

Value

sorted data.frame of CI data.

Examples

example(savR)
colnames(correctedIntensities(fc))
cycles \hspace{2cm} \textit{Get the total number of cycles}

\textbf{Description}
Accessor to obtain the total number of cycles sequenced in an Illumina sequencing run.

\textbf{Usage}
\begin{verbatim}
cycles(project)

## S4 method for signature 'savProject'
cycles(project)
\end{verbatim}

\textbf{Arguments}
\begin{verbatim}
project \hspace{2cm} SAV project
\end{verbatim}

\textbf{Value}
total number of cycles in run, including all sequencing and index reads.

\textbf{Examples}
\begin{verbatim}
extample(savR)
cycles(fc)
\end{verbatim}

directions \hspace{2cm} \textit{Get the number of sequence reads}

\textbf{Description}
Returns the number of sequencing reads (excluding index reads).

\textbf{Usage}
\begin{verbatim}
directions(project)

## S4 method for signature 'savProject'
directions(project)
\end{verbatim}

\textbf{Arguments}
\begin{verbatim}
project \hspace{2cm} SAV project
\end{verbatim}

\textbf{Value}
number of reads

\textbf{Examples}
\begin{verbatim}
extample(savR)
directions(fc)
\end{verbatim}
errorMetrics

Description
Error metrics for lane, tile, and cycle.

Usage
```r
errorMetrics(project)
```

Arguments
- `project` SAV project

Details
- `lane`: Lane number
- `tile`: Tile ID
- `cycle`: Cycle number
- `errorrate`: Error rate
- `nPerfect`: number of perfect reads
- `n[1-4]Error`: Number of reads with 1, 2, 3 and 4 errors

Value
sorted data.frame of Error metrics

Examples
```r
example(savR)
colnames(extractionMetrics(fc))
```

extractionMetrics

Description
Extraction (intensity and FWHM) metrics for lane, tile, and cycle.

Usage
```r
extractionMetrics(project)
```

Arguments
- `project` SAV project

Details

Value

Examples
```r
example(savR)
colnames(extractionMetrics(fc))
```
Arguments

project  SAV project

Details

lane: Lane number
tile: Tile ID
cycle: Cycle number
FWHM_[ACGT]: Full width at half maximum for A, C, G, or T
int_[ACGT]: Intensity of channel A, C, G, or T
datestamp: Time/date stamp

Value

sorted data.frame of Extraction metrics

Examples

example(savR)
colnames(extractionMetrics(fc))

---

flowcellLayout  Get flowcell layout

Description

Accessor to obtain information about the characteristics of the flowcell from an Illumina sequencing run.

Usage

flowcellLayout(project)

## S4 method for signature 'savProject'
flowcellLayout(project)

Arguments

project  SAV project

Value

illuminaFlowCellLayout-class object

Examples

eexample(savR)
flowcellLayout(fc)
**illuminaFlowCellLayout-class**

*Layout of an Illumina flowcell*

**Description**

Class representation of the features of an Illumina flow cell.

**Slots**

- `lanecount`: Number of lanes on the flowcell
- `surfacecount`: Number of surfaces
- `swathcount`: Number of imaging swaths
- `tilecount`: Number of tiles per swath
- `sectionperlane`: Number of sections per lane (NextSeq)
- `lanepersection`: Number of lanes per section (NextSeq)
- `tilenamingconvention`: Description of deviation from original formatting layout

**illuminaRead-class**

*Illumina read*

**Description**

Class representation of the features of an Illumina sequencing read.

**Slots**

- `number`: the index of this read in sequencing
- `cycles`: number of cycles in this read
- `index`: logical representing whether or not this read is an index read

**location**

*Get Flowcell folder location*

**Description**

_accessor to obtain the path to data for a particular SAV project.*

**Usage**

location(project)

## S4 method for signature 'savProject'

location(project)
pfClusters

**Arguments**

- `project` SAV project

**Value**

normalized path to Illumina run data.

**Examples**

- `example(savR)`
- `location(fc)`

---

pfBoxplot  
**PF Boxplot**

**Description**

Generate a boxplot of the numbers of clusters and the number of Illumina pass-filter clusters per tile and lane.

**Usage**

```r
pfBoxplot(project)
## S4 method for signature 'savProject'
pfBoxplot(project)
```

**Arguments**

- `project` SAV project

---

pfClusters  
**Get number of PF clusters per lane**

**Description**

Sum the total pass filter number of clusters for all tiles in the lane.

**Usage**

```r
pfClusters(project, lane)
## S4 method for signature 'savProject,integer'
pfClusters(project, lane = 1L)
```

**Arguments**

- `project` SAV project
- `lane` lane(s) number
plotFWHM

Examples

```r
## Not run:
example(savR)
pfClusters(fc, 1L)

## End(Not run)
```

---

### Generate FWHM plots

**Description**

Plots the average full width of clusters at half maximum (FWHM) of each tile for a given cycle and base.

**Usage**

```r
plotFWHM(project, cycle, base)
```

**Arguments**

- `project`: SAV project
- `cycle`: sequence cycle
- `base`: nucleotide base (ACGT)

---

plotIntensity

**Plot flowcell intensity by base and cycle**

**Description**

Draws a representation of a flowcell, showing the average corrected called intensity values.
Usage

plotIntensity(project, cycle, base)

## S4 method for signature 'savProject, integer, character'
plotIntensity(project, cycle = 1L,
base = c("A", "C", "G", "T"))

## S4 method for signature 'savProject, missing, missing'
plotIntensity(project)

## S4 method for signature 'savProject, integer, missing'
plotIntensity(project, cycle)

## S4 method for signature 'savProject, missing, character'
plotIntensity(project, base)

Arguments

- project: A `savProject`-class object
- cycle: integer cycle number
- base: character for nucleotide

---

**plotQGT30**

*Plot Quality > 30 for a flowcell*

Description

Generate a plot for a given cycle of the percentage of clusters in each tile that are \( \geq \) Q30.

Usage

plotQGT30(project, cycle)

## S4 method for signature 'savProject, integer'
plotQGT30(project, cycle = 1L)

## S4 method for signature 'savProject, missing'
plotQGT30(project)

Arguments

- project: SAV project
- cycle: sequence cycle
qualityHeatmap

Generate a heatmap of qualities

Description

Plots a heatmap of quality vs cycle for a given lane for 1 or more sequence reads. Read qualities include sequence + index.

Usage

```
qualityHeatmap(project, lane, read, collapse)
```

```
## S4 method for signature 'savProject,integer,integer,logical'
qualityHeatmap(project, lane,
    read, collapse = T)
```

```
## S4 method for signature 'savProject,numeric,numeric,missing'
qualityHeatmap(project, lane,
    read)
```

Arguments

- **project**: SAV project
- **lane**: integer lane specification
- **read**: integer vector of sequence reads to include (not including index reads)
- **collapse**: whether or not to collapse index reads into the preceding read (# reads = directions), default TRUE

qualityMetrics

Get Quality Metrics data

Description

Quality metric by lane, tile and cycle.

Usage

```
qualityMetrics(project)
```

```
## S4 method for signature 'savProject'
qualityMetrics(project)
```

Arguments

- **project**: SAV project
Details

lane: Lane number
tile: Tile ID
cycle: Cycle number
Q1–Q50: Number of clusters with quality of indicated column

Value

sorted data.frame of quality data

Examples

example(savR)
colnames(qualityMetrics(fc))

Description

Accessor to obtain information about the reads of a particular Illumina sequencing run.

Usage

reads(project)

## S4 method for signature 'savProject'
reads(project)

Arguments

project SAV project

Value

List of illuminaRead-class objects

Examples

example(savR)
reads(fc)
run

Get the Run ID

Description
Accessor to obtain the string identifier of an Illumina sequencing run.

Usage
run(project)

## S4 method for signature 'savProject'
run(project)

Arguments
project SAV project

Value
parsed Illumina run id

Examples
example(savR)
run(fc)

savCorrectedIntensityFormat-class
Corrected Intensity formatter

Description
Lane, tile, cycle, average intensity, corrected intensities (ACGT), average corrected called intensities (ACGT), number of no-calls, number of (ACGT) calls, and signal to noise ratio.

Slots
name: vector of column names
type: vector of data types of elements
lengths: vector of byte lengths for each element
order: vector of column names for sorting
version: integer version number
savData-class

Structure for holding parsed InterOp headers and data

Description
Structure for holding parsed InterOp headers and data

Slots
header: list of parsed header values
data: data.frame of parsed values

savErrorFormat-class
Error Metrics formatter

Description
Lane, tile, cycle, errorrate, nPerfect, n1Error, n2Error, n3Error, n4Error.

Slots
name: vector of column names
type: vector of data types of elements
lengths: vector of byte lengths for each element
order: vector of column names for sorting
version: integer version number

savExtractionFormat-class
Extraction Metrics formatter

Description
Lane, tile, cycle, FWHM (ACGT), intensity (ACGT), datestamp, timestamp. Datestamp and timestamp are munged at the moment because R does not have native support for 32-bit unsigned integers and I have not implemented a solution.

Slots
name: vector of column names
type: vector of data types of elements
lengths: vector of byte lengths for each element
order: vector of column names for sorting
version: integer version number
savFormat-class

**Base class for formatters**

**Description**

Defines the necessary slots to create parse different binary files using the same generic parser.

**Slots**

- name: vector of column names
- type: vector of data types of elements
- lengths: vector of byte lengths for each element
- order: vector of column names for sorting
- version: integer version number
- default: logical default format ()

savProject-class

**SAV project class**

**Description**

Represents a flowcell, metadata and parsed SAV information

**Slots**

- location: Full path to flowcell directory
- reads: List of illuminaRead-class
- layout: illuminaFlowCellLayout-class
- runid: Run ID
- number: Run number
- flowcell: Flowcell ID
- instrument: Instrument ID
- date: Run date
- cycles: Total number of cycles
- directions: Total number of sequence runs (ends)
- parsedData: SAV data
**savQualityFormat-class**

*Quality Metrics formatter*

**Description**

Lane, tile, cycle, Q1-Q50 counts

**Slots**

- **name**: vector of column names
- **type**: vector of data types of elements
- **lengths**: vector of byte lengths for each element
- **order**: vector of column names for sorting
- **version**: integer version number

**savQualityFormatV5-class**

*Quality Metrics formatter version 5*

**Description**

Lane, tile, cycle, Q1-Q50 counts

**Slots**

- **name**: vector of column names
- **type**: vector of data types of elements
- **lengths**: vector of byte lengths for each element
- **order**: vector of column names for sorting
- **version**: integer version number

**savR**

*Build a SAV project*

**Description**

Constructor to build a **savProject-class** object and populate it. A SAV project consists of binary files generated by an Illumina sequencing run and placed in a folder named "InterOp". This folder contains a number of ".bin" files that contain statistics about the run. Creating this object parses all of the files and makes the data available for analysis.
savTileFormat-class

Usage

savR(object)

## S4 method for signature 'character'
savR(object)

## S4 method for signature 'missing'
savR()

Arguments

object String Path to Flowcell data

Examples

fc <- savR(system.file("extdata", "MiSeq", package="savR"))
fc

savTileFormat-class  Tile Metrics formatter

Description

Lane, tile, code, value. Codes are:

Details

100  Cluster Density
101  PF Cluster Density
102  Number of clusters
103  Number of PF clusters
400  Control lane

Slots

name: vector of column names
type: vector of data types of elements
lengths: vector of byte lengths for each element
order: vector of column names for sorting
version: integer version number (header consists of version (1b), length (1b))
tileMetrics  

Description

Returns the Tile Metrics SAV data.

Usage

```r
tileMetrics(project)
```

## S4 method for signature 'savProject'
tileMetrics(project)

Arguments

- `project` SAV project

Details

Metrics for each tile are encoded in the following format:

- cluster density: 100
- PF cluster density: 101
- number of clusters: 102
- number of PF clusters: 103
- phasing for read N: \((200 + (N - 1) \times 2)\)
- prephasing for read N: \((201 + (N - 1) \times 2)\)
- percent aligned for read N: \((300 + N - 1)\)
- control lane: 400

- lane: Lane number
- tile: Tile ID
- code: Code described above
- value: Value for code key

Value

- sorted data.frame of tile metrics

References

Codes for Tile Metrics were obtained from the Python Illuminate package:

[https://bitbucket.org/invitae/illuminate](https://bitbucket.org/invitae/illuminate)

Examples

```r
example(savR)
colnames(tileMetrics(fc))
```
Index

*Topic package
  savR-package, 2
buildReports, 3
buildReports, savProject, character-method (buildReports), 3
buildReports, savProject, missing-method (buildReports), 3
clusterQualityGtN, 3
clusterQualityGtN, savProject, integer, integer, integer (clusterQualityGtN), 3
clusterQualityGtN, savProject, integer, integer, integer-method (clusterQualityGtN), 3
clusters, 4
clusters, savProject, integer-method (clusters), 4
correctedIntensities, 5
correctedIntensities, savProject-method (correctedIntensities), 5
cycles, 6
cycles, savProject-method (cycles), 6
directions, 6
directions, savProject-method (directions), 6
errorMetrics, 7
errorMetrics, savProject-method (errorMetrics), 7
extractionMetrics, 7
extractionMetrics, savProject-method (extractionMetrics), 7
flowcellLayout, 8
flowcellLayout, savProject-method (flowcellLayout), 8
illuminaFlowCellLayout-class, 8, 9, 17
illuminaRead-class, 9, 14, 17
location, 9
location, savProject-method (location), 9
pfBoxplot, 10
pfBoxplot, savProject-method (pfBoxplot), 10
pfClusters, 10
pfClusters, savProject, integer (pfClusters), 10
pfClusters, savProject, integer-method (pfClusters), 10
plotFWHM, 11
plotFWHM, savProject, integer, character-method (plotFWHM), 11
plotFWHM, savProject, integer, missing-method (plotFWHM), 11
plotFWHM, savProject, missing, character-method (plotFWHM), 11
plotFWHM, savProject, missing, missing-method (plotFWHM), 11
plotIntensity, 11
plotIntensity, savProject, integer, character-method (plotIntensity), 11
plotIntensity, savProject, integer, missing-method (plotIntensity), 11
plotIntensity, savProject, missing, character-method (plotIntensity), 11
plotIntensity, savProject, missing, missing-method (plotIntensity), 11
plotQGT30, 12
plotQGT30, savProject, integer-method (plotQGT30), 12
plotQGT30, savProject, missing-method (plotQGT30), 12
qualityHeatmap, 13
qualityHeatmap, savProject, integer, integer, logical-method (qualityHeatmap), 13
qualityHeatmap, savProject, numeric, numeric, missing-method (qualityHeatmap), 13
qualityMetrics, 13
qualityMetrics, savProject-method (qualityMetrics), 13
reads, 14
reads, savProject-method (reads), 14
run, 15
run, savProject-method (run), 15
savCorrectedIntensityFormat-class, 15
savData-class, 16
savErrorFormat-class, 16
savExtractionFormat-class, 16
savFormat-class, 17
savProject-class, 12, 17, 18
savQualityFormat-class, 18
savQualityFormatV5-class, 18
savR, 18
savR, character-method (savR), 18
savR, missing-method (savR), 18
savR-package, 2
savTileFormat-class, 19

tileMetrics, 20
tileMetrics, savProject-method (tileMetrics), 20