Creating IGV HTML reports with tracktables

Thomas Carroll[1*]

[1] Bioinformatics Facility, MRC Clincal Sciences Centre;

[*]`thomas.carroll (at)imperial.ac.uk`

September 22, 2014

Abstract

The visualisation of genomics data in genome browsers is a key step in both quality control and initial interrogation of hypothesis under investigation.

The organisation of large collections of genomics data (such as large scale High-thoughput sequencing data experiments) alongside assocated Sample or Experimental metadata allows for rapid evaluation of patterns across such experimental groups but such organisation maybe very time consuming.

The Tracktables package provides a set of tools to build IGV session files from sample files and associated metadata as well as produce IGV linked HTML reports for high thoroughput visualisation of sample data in IGV.

Contents

1   Creating IGV sessions and HTML reports using tracktables

The three main functions within tracktables package are the MakeIGVSessionXML() and MakeIGVSampleMetadata() functions for a creating IGV session for sample files and sample metadata and the maketracktable() function to create HTML pages with tables to control IGV alongside sample information and metadata,

1.1   Creating input files for tracktables

Tracktables functions require the user to provide both a matrix or dataframe of metadata information and one of sample file locations to be visualised in IGV.

These matricies/data-frames must both have one column names "SampleSheet" which contains unique sample IDs.

The remaining metadata samplesheet columns may be user-defined but must all contain column titles. (See example below) The sample filesheet must contain the columns "SampleName", "bam","bigwig" and interval. These columns may contain NA values when no relavant file is associated to a sample.

Here we create a small example SampleSheet (containing metadata) and FileSheet (containing file locations)

```r
library(tracktables)

fileLocations <- system.file("extdata",package="tracktables")

bigwigs <- dir(fileLocations,pattern="*.bw",full.names=T)

intervals <- dir(fileLocations,pattern="*.bed",full.names=T)

bigWigMat <- cbind(gsub("_Example.bw","",basename(bigwigs)),
                   bigwigs)

intervalsMat <- cbind(gsub("_Peaks.bed","",basename(intervals)),
                      intervals)

fileSheet <- merge(bigWigMat,intervalsMat,all=T)
```

```
fileSheet <- as.matrix(cbind(fileSheet,NA))

colnames(fileSheet) <- c("SampleName","bigwig","interval","bam")

SampleSheet <- cbind(as.vector(fileSheet[,"SampleName"]),
                     c("EBF","H3K4me3","H3K9ac","RNAPol2"),
                     c("ProB","ProB","ProB","ProB"))

colnames(SampleSheet) <- c("SampleName","Antibody","Species")
```

```
head(SampleSheet)
```

```
##      SampleName Antibody  Species
## [1,] "EBF"      "EBF"     "ProB"
## [2,] "H3K4me3"  "H3K4me3" "ProB"
## [3,] "H3K9ac"   "H3K9ac"  "ProB"
## [4,] "RNAPol2"  "RNAPol2" "ProB"
```

The SampleSheet can be seen to contain a very small section of metadata for EBF, RNApol2, H3K4me3 and H3K9ac. SampleName column contains the unique IDs.

```
head(fileSheet)
```

The FileSheet contains the "SampleName" column with unique ID which match with those found in the SampleSheet. The remaining columns of "bam","bigwig" and "interval" are also all present.

Note that not all samples have intervals associated to them and ,here, none of these samples have BAM files associated to them. NA values within the fileSheet wil be ignored by tracktables.

1.2   Creating an IGV session XML file

Tracktables can create an IGV session XML and associated sample information file from this samplesheet and filesheet.

In addition to the fileSheet and samplesheet, the MakeIGVSampleMetadata and MakeIGVSessionXML() functions require the location to write to, filename for XML and the genome to be used in IGV (see IGV for detials on supported genomes).

```
MakeIGVSampleMetadata(SampleSheet,fileSheet,igvdirectory=getwd())
MakeIGVSessionXML(fileSheet,igvdirectory=getwd(),"Example","mm9")
```

This creates two files in the current working directory containing the sample information file for IGV "SampleMetadata.txt" and the session itself to be loaded into IGV "Example.xml".

1.3 Creating a Tracktable report HTML page

As well as producing standalone session XMLs and sampleinformation files, tracktables package can produce HTML reports which contain metadata and link to control IGV.

The report structure is made of a main "Tracktables Sample Report" which houses the metadata from "SampleSheet" and links to open sample files in IGV. All sample files are associated with their relevant sample metadata and grouped together by their unique SampleName. When a sample has an interval files associated to it, the "Tracktables Sample Report" also contains a link to a further sample specific "Tracktables Interval Report". This interval report contains a table of interval locations, any metadata associated with intervals and further links to focus IGV on that Intervals region.

```
HTMLreport <- maketracktable(fileSheet,SampleSheet,
                             "IGV_Example",
                             basedirectory=getwd(),
                             "mm9")
```