Genomic Data Analysis in R

Lecture 8

October 22, 2019

Gavin Ha, Ph.D.

Assistant Member

Public Health Sciences

FRED HUTCH
CURES START HERE®

Overview

I. R and Bioconductor Packages for Genomic Data

II. Tutorials

- **Analyzing Sequence Data**

- **Analyzing Genomic Variants**

- **Annotating Variants**

FRED HUTCH

Overview: Learning Objectives

I. R Bioconductor Packages for Genomic Data

- `Rsamtools, GenomeInfoDb, GenomicRanges, VariantAnnotation`

II. Tutorials

   1. Sequence Data Analysis

      i. Load, inspect, query a BAM alignment file

      ii. Extract sequences and qualities

      iii. Compute "pile-up" statistics at genomic loci

   2. Genomic Data Analysis

      i. Load, inspect, query a BED/MAF file

      ii. Load, inspect, query a VCF file

   3. Genomic Variant Annotations

FRED HUTCH

Tutorial #1: Sequence Data Analysis

1. Loading and querying a BAM file using Rsamtools

    a. Define the genomic coordinates and components to query (`ScanBamParam`)

    b. Scanning the BAM file (`scanBam`)

2. We will use the example from Lecture 7: Slides 12-16.

3. Download the BAM file for this tutorial

    • https://www.dropbox.com/sh/zoitjnobgp7l7c2/AABBIpTQcNA4IWYOFnV5dlMKa?dl=0

    • BRCA_IDC_cfDNA.bam, BRCA_IDC_cfDNA.bam.bai

4. R Markdown file for tutorial on GitHub: `Lecture8_Rsamtools.Rmd`

Tutorial #2: Genomic Data Analysis

1. Loading and querying BED/VCF text files

   a. Use packages `GenomicRanges, VariantAnnotation, data.table`

2. Concepts from Lecture 7: Slides 38-42.

3. Download the BAM file for this tutorial

   - https://www.dropbox.com/sh/zoitjnobgp7I7c2/AABBIpTQcNA4IWYOFnV5dIMKa?dl=0

   - BRCA.genome_wide_snp_6_broad_Level_3_scna.seg.bed

   - GIAB_highconf_v.3.3.2.vcfgz, GIAB_highconf_v.3.3.2.vcf.gz.tbi

4. R Markdown file for tutorial on GitHub: `Lecture8_GenomicData.Rmd`

Tutorial #3: Genomic Variant Annotations

1. Loading and querying UCSC and Ensemble transcript/gene databases in R

   a. Use packages `biomaRt` (Ensembl),
      `TxDb.Hsapiens.UCSC.hg19.knownGene`

2. Download the BAM file for this tutorial

   • https://www.dropbox.com/sh/zoitjnobgp7l7c2/AABBIpTQcNA4IWYOFnV5dIMKa?dl=0

   • BRCA.genome_wide_snp_6_broad_Level_3_scna.seg.bed

   • GIAB_highconf_v.3.3.2.vcfgz, GIAB_highconf_v.3.3.2.vcf.gz.tbi

3. R Markdown file for tutorial on GitHub: `Lecture8_Annotations.Rmd`