AnnotateGenomicRegions: a Web application

L. Zammataro, G. Bucci, H. Muller✉

Computational Research, Center for Genomic Science of IIT@SEMM c/o IFOM-IEO-CAMPUS (Italian Institute of Technology), Milano, Italy

Motivations

Next-generation sequencing (NGS) is producing large data volumes at reasonable cost and new applications are being developed at increasing speed. A common denominator for all applications of NGS technology is the need to annotate genomic regions of interest. Tools such as Galaxy [1], CisGenome [2], or the Bioconductor ChIPpeakAnno package [3] have been published to perform this task. However, using these tools often requires a significant amount of bioinformatics skills and/or downloading and installing dedicated software. A widely accepted, web-based annotation tool available to bioinformaticians and biologists with widely varying skill levels is not available. Indeed, many skilled bioinformaticians rely on self-made scripts to process the data to be annotated in the desired input/output format and in the necessary detail. For many biologists working with new generation sequencing data, annotating a set of genomic regions represents a complicated task that necessarily involves the help of a skilled bioinformatician.

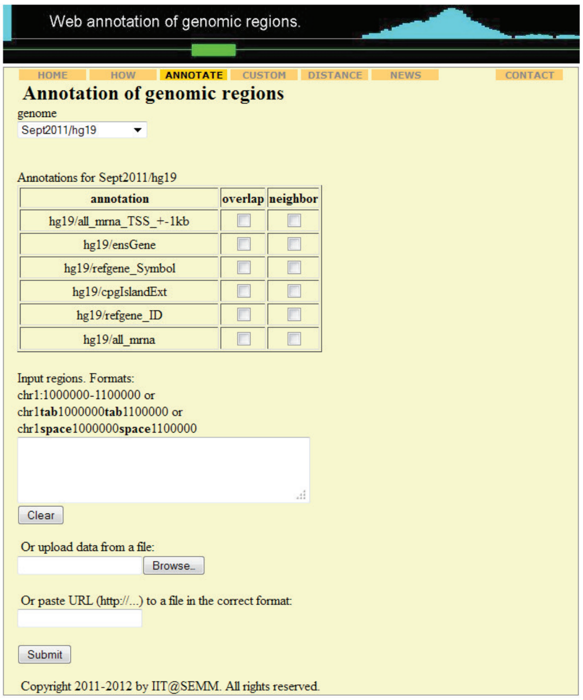
Methods

Here we present AnnotateGenomicRegions, a web application that accepts genomic regions as input and outputs overlapping and/or neighboring genome annotations chosen on a simple web-form. The application is based on Java Enterprise technology and runs on a Glassfish server. The necessary speed of annotating hundreds of thousands of genomic regions with tens of different annotations within seconds is achieved using a proprietary hash-based data structure.

Results

We developed an annotation tool that fulfills five basic design criteria:

1. genomic regions shall be used as input query;
2. the output shall be pastable into an Excel table;
3. the application shall be web-based;
4. no programming skills required to use the application;



Annotation of genomic regions

genome
Sept2011/hg19

Annotations for Sept2011/hg19

annotation	overlap	neighbor
hg19/all_mrna_TSS_+-1kb	☐	☐
hg19/ensGene	☐	☐
hg19/refgene_Symbol	☐	☐
hg19/cpgIslandExt	☐	☐
hg19/refgene_ID	☐	☐
hg19/all_mrna	☐	☐

Input regions. Formats:
chr1:1000000-1100000 or
chr1**tab**1000000**tab**1100000 or
chr1**space**1000000**space**1100000

Or upload data from a file:

Or paste URL (http://...) to a file in the correct format:

Copyright 2011-2012 by IIT@SEMM. All rights reserved.

Figure 1.

5. it must be fast enough to annotate hundreds of thousands of genomic regions within seconds.

The tool can be installed on any computer capable of running Java and Glassfish on a Windows or Unix/Linux operating system, which is from a laptop to a mainframe computer.

Availability

http://sourceforge.net/projects/annotatelocus/?source=directory

References

1. Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, et al. Galaxy: a platform for interactive large-scale genome analysis. Genome Res. 2005;15:1451-5.
2. Ji H, Jiang H, Ma W, Johnson DS, Myers RM, Wong WH. An integrated software system for analyzing ChIP-chip and ChIP-seq data. Nat Biotechnol. 2008;26:1293-300.
3. Zhu LJ, Gazin C, Lawson ND, Pages H, Lin SM, Lapointe DS, et al. ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC Bioinformatics. 2010;11:237.