CHROMATRA: a Galaxy tool for visualizing genome-wide chromatin signatures

Thomas Hentrich1,2, Julia M. Schulze2, Eldon Emberly3,∗ and Michael S. Kobor2,∗

1Department of Computing Science, University of British Columbia, Vancouver, BC V6T 1Z4, 2Department of Medical Genetics, Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, Vancouver, BC V5Z 4H4 and 3Department of Physics, Simon Fraser University, Burnaby, BC V5A 1S6, Canada

ABSTRACT

CHROMATRA (CHROmatin Mapping Across TRAnscripts) is a visualization tool available as plug-in for the Galaxy platform. It allows detailed yet concise presentations of data derived from ChIP-chip or ChIP-seq experiments by visualizing enrichment scores across genes or other genomic features while accounting for their length and additional characteristics such as gene expression. It integrates into typical analysis workflows and enables rapid graphical assessment and comparison of genome-wide data at a glance.

Availability: https://github.com/cmmt/chromatra

Received on September 22, 2011; revised on November 30, 2011; accepted on January 3, 2012

1 INTRODUCTION

The pace of advancements in DNA microarray (Stoughton, 2005) and sequencing technology (Shendure and Ji, 2008) continues to transform biological research. Both techniques complement each other and provide powerful platforms to interrogate entire genomes at high resolution, ranging from sequence decoding, to gene expression profiling, to chromatin modification mapping (Collas, 2010; Kapranov et al., 2007; Metzker, 2010).

It is in large part due to these technologies that data are becoming available to scientists and allow questions to be asked of unprecedented depth. At the same time, however, the increasingly large volume and complexity of datasets demands adequate computational methods to analyze and present the data effectively.

While raw numbers may contain all relevant information of an analysis, they are not always necessarily suitable to convey biological meaning efficiently, especially when presenting data in scientific publications. Graphic visualizations of data can aid in representing complex content and reveal salient information at a glance (Gehlenborg et al., 2010). Yet, satisfying this claim is particularly challenging for genome-wide data. For instance, when mapping chromatin modifications or transcription factors using ChIP-chip or ChIP-seq, it is a major goal to assess their localization and distribution across all genes of an organism while preserving structural and functional features of their respective genome.

Current approaches for visualizing binding or modification profiles can be broadly distinguished into two categories: first, tools of single-gene resolution, like the Genome Browser (Fujita et al., 2011), offer the most detailed level of data presentation, but make it difficult to evaluate profiles across genes and derive broader principles as they disperse the information over multiple plots to cover the genome. Second, averaging approaches try to circumvent this problem by grouping genes into classes, e.g. based on length, Gene Ontology terms (Ashburner et al., 2000) or transcription rate (Mayer et al., 2010; Pokholok et al., 2005) and calculating average profiles for each class. While they succeed in condensing the mosaic representation of single-gene visualizations, they inevitably sacrifice detail dependent on the applied grouping given the heterogeneity of genes with respect to their length and other features.

Here, we present CHROMATRA (CHROmatin Mapping Across TRAnscripts), a visualization tool for genome-wide DNA-protein binding and chromatin modification maps that balances detail and compactness of the existing visualization approaches. Developed as a plug-in for the Galaxy analysis environment (Goecks et al., 2010), CHROMATRA portrays enrichment profiles in a comprehensive yet condensed form by accounting for feature length and additional characteristics such as gene expression. CHROMATRA plots reveal the spatial distribution of chromatin marks or binding events across all genomic features of interest, and may aid in discovering unexpected correlations and patterns in enrichment profiles.

2 FUNCTIONALITY AND WORKFLOW

CHROMATRA is a plug-in for the Galaxy bioinformatics platform (Goecks et al., 2010). As one of the most comprehensive analysis solutions for genomics data, Galaxy integrates different software tools under a unified user interface with advanced workflow management capabilities (Blankenberg et al., 2007). Since CHROMATRA was written in Python it easily integrates with Galaxy, itself implemented in Python, and has no dependencies on external libraries not already used by Galaxy.

CHROMATRA consists of two visualization modules, which can be integrated into the analysis workflow for ChIP-chip or ChIP-seq experiments. Typically, the raw data of ChIP experiments are normalized and enrichment scores calculated and stored in standardized formats such as WIG, BED and GFF for subsequent analysis. After these steps, CHROMATRA can be linked to the workflow and used to visualize ChIP profiles by reading GFF files that contain enrichment scores for the whole genome.

The first module, CHROMATRA-L, accounts for differences in lengths of genomic features, such as transcripts, exons or telomeric

∗To whom correspondence should be addressed.
T.Hentrich et al.

Fig. 1. CHROMATRA plot based on ChIP-chip profile of histone variant H2A.Z in S. cerevisiae (data from Wang et al., 2009). (A) CHROMATRA-L plot for H2A.Z across transcripts in yeast, sorted by length and aligned by TSS, showing that H2A.Z is enriched primarily in promoters but absent in transcript bodies. Normalized enrichment scores were binned into segments of 150 bp. (B) CHROMATRA-T plot for same dataset as in (A), showing that H2A.Z preferentially marks infrequently transcribed genes and is absent in transcripts with high transcription rate. Transcripts sorted by length and transcription rate range (Helsteg et al., 1998).

3 CONCLUSION AND OUTLOOK

CHROMATRA implements two visualization approaches for exploring genome-wide chromatin signatures and facilitating the discovery of unexpected patterns and correlations. The comprehensive yet compact visualization allows assessing enrichment profiles for data from any organism. The CHROMATRA modules have proven their effectiveness in a recent study, addressing the localization and dependencies of chromatin modifications in the context of transcription (Schulze et al., 2011). Besides assessing genes or transcripts, any other genomic feature that can be aligned according to a common position can be visualized, including introns and alternative transcripts. Furthermore, data derived through related techniques, such as MeDIP-seq (Jacinto et al., 2008; Weber et al., 2005) or MBD-seq (Serre et al., 2010), can be used as input as well. As a plug-in for Galaxy, CHROMATRA easily integrates into existing workflows and enables direct graphical exploration of genome-wide data.

ACKNOWLEDGEMENTS

We thank Dr Arvind Gupta (UBC) for support and helpful discussions during the development of this project. TH is a fellow of the CIHR/MSHR Bioinformatics Training Program.

Funding: Canadian Institute of Health Research (MOP-79442).

Conflict of Interest: none declared.

REFERENCES