I2EHR: an interactive dashboard for integrative analysis of clinical and genomic data

Abstract

Summary: I2EHR is an approach to analysing integrated clinical and genomic data and identifying patterns that are not commonly identifiable in unstructured clinical data. A lack of standardisation has hindered the integration of clinical and genomic data and limited the potential for patient and cohort level analysis. We present I2EHR, a dashboard for integrative analysis for detection of clinically relevant disease patterns at the patient and cohort level. I2EHR increases the usefulness of longitudinal wellness measurements with visualisations of variation over time and in relation to gene expression or other clinical measurements. By displaying differential expression analysis results along clinical data, new cohort subgroups can be identified and used for clinical decision support.

Availability and implementation: I2EHR is implemented as an R Shiny package and is publically available at https://github.com/shanecrinion/I2EHR

Contact: shanecrinion@gmail.com

Introduction

Differential signal detection in sequencing data is one of the most common tasks in genomic analyses. Multiple tools have been developed for this purpose, many of which, including DESeq and EdgeR, are based on the negative binomial models for count data (Anders and Huber, 2010; Robinson et al., 2010). Such tools are theoretically suitable for the analysis of most sequencing data types, including chromatin immunoprecipitation (ChIP-seq) and Hi-C, leading to the development of wrapper packages around DESeq and EdgeR that facilitate differential analyses for such data (Ross-Innes et al., 2012; Lareau and Aryee, 2018). However, both of these algorithms have been developed with standard RNA sequencing data in mind and therefore not account for or benefit from the specific properties of data resulting from other assays, prompting the development of assay-specific differential analysis tools (Xu et al., 2008; Chen et al., 2015; Stansfield et al., 2018; Liu and Ruan, 2017)

Capture Hi-C (CHi-C) is a powerful experimental technique for detecting chromosomal interactions globally and at high resolution (Schoenfelder et al., 2015). In CHi-C, the genome-wide pulldown of pairs of interacting genomic fragments by Hi-C is followed by sequence capture to selectively enrich Hi-C material for interactions involving (at least on one end) fragments of interest, termed 'baits'. Differential analyses of CHi-C data are challenging due to sample normalisation issues, sparsity and uneven signal-to-noise ratios across interaction distances and different capture baits, which are not accounted for by standard differential analysis algorithms. We have previously reported CHiCAGO, a statistical pipeline for

robust detection of significant interactions in Capture Hi-C data from a single condition (Cairns et al., 2016). Here, we present Chicdiff, an R package for differential Capture Hi-C data analysis. Chicdiff combines moderated differential testing for count data implemented in DESeq2 (Love et al., 2014) with CHi-C-specific procedures for signal normalisation informed by CHiCAGO and p-value weighting. Jointly, procedures

implemented in Chicdiff enable a robust and sensitive detection of differential interactions in CHi-C data.