Exploring Lineage-Specific Enhancers by Integrating Enhancer Transcription, Epigenomic Features, Sequence Motifs, and Transcription Factor Expression

This manuscript was automatically generated from vsmalladi/tfsee-manuscript@9cddd2c on November 4, 2017.

Authors

- **Venkat Malladi**
  0000-0002-0144-0564 · vsmalladi · katatonikkat

Abstract

The identification of transcription factors (TF) driving the formation of active enhancers that regulate the expression of target genes remains an open problem. We have developed a computational framework that identifies cell type-specific enhancers and their cognate TFs by integrating multiple genomic assays that probe the transcriptomes (GRO-seq and RNA-seq) and epigenomes (ChIP-seq) of various samples. Our method, called Total Functional Score of Enhancer Elements (TFSEE), integrates the magnitude of enhancer transcription (GRO-seq), enrichment of marks associated with enhancers (H3K4me1 and H3K27ac ChIP-seq), TF mRNA expression levels (RNA-seq), and TF motif p-values (MEME). This method has allowed us to explore the enhancer landscape in different cell types that share common origins or are biologically related, including distinct molecular subtypes of breast cancer, and embryonic stem cells (ESCs) and their derived lineages. Using TFSEE, we have identified key breast cancer subtype-specific transcription factors that are bound at active enhancers and dictate gene expression patterns determining growth outcomes. To demonstrate the broader utility of our approach, we have used this algorithm to identify transcription factors during the differentiation of embryonic stem cells into pancreatic cells. Taken together our results show that TFSEE can be used to perform multilayer genomic data integration to uncover novel cell type-specific transcription factors that control lineage-specific enhancers.

Introduction

Results

Discussion

Acknowledgments

Material and Methods

Genomic Data Curation

We used previously published GRO-seq, ChIP-seq and RNA-seq data from [1,2] of time course differentiation of human embryonic stem cells (hESC) to pancreatic endoderm (PE). All data sets are available from NCBI's Gene Expression Omnibus repository [3] or EMBL-EBI's ArrayExpress repository [4] using the accession numbers listed in Table 1.

Table 1: **Description and accession numbers of GRO-seq, ChIP-seq and RNA-seq datasets.**

Assay	Accessions
GRO-seq	GSM1316306, GSM1316313, GSM1316320, GSM1316327, GSM1316334
H3K4me3 ChIP-seq	ERR208008, ERR208014, ERR207998, ERR20798, ERR207999
H3K4me1 ChIP-seq	GSM1316302, GSM1316303, GSM1316309, GSM1316316, GSM1316317, GSM1316310, GSM1316323, GSM1316324, GSM1316330, GSM1316331
H3K27ac ChIP-seq	GSM1316300, GSM1316301, GSM1316307, GSM1316308, GSM1316314, GSM1316315, GSM1316321, GSM1316322, GSM1316328, GSM1316329

Assay	Accessions
Input ChIP-seq	ERR208001, ERR208012, ERR207984, ERR208011, ERR207986, GSM1316304, GSM1316305, GSM1316311, GSM1316312, GSM1316318, GSM1316319, GSM1316325, GSM1316326, GSM1316332, GSM1316333
RNA-seq	ERR266333, ERR266335, ERR266337, ERR266338, ERR266341, ERR266342, ERR266344, ERR266346, ERR266349, ERR266351

References

1. **Dynamic chromatin remodeling mediated by Polycomb proteins orchestrates pancreatic differentiation of human embryonic stem cells**
Ruiyu Xie, Logan J. Everett, Hee-Woong Lim, Nisha A. Patel, Jonathan Schug, Evert Kroon, Olivia G Kelly, Allen Wang, Kevin A. D'Amour, Allan J. Robins, … Maike Sander
Cell stem cell (2013-02-07) https://www.ncbi.nlm.nih.gov/pubmed/23318056

2. **Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates**
Allen Wang, Feng Yue, Yan Li, Ruiyu Xie, Thomas Harper, Nisha A. Patel, Kayla Muth, Jeffrey Palmer, Yunjiang Qiu, Jinzhao Wang, … Maike Sander
Cell stem cell (2015-04-02) https://www.ncbi.nlm.nih.gov/pubmed/25842977

3. **GEO**
Gene Expression Omnibus
https://www.ncbi.nlm.nih.gov/geo/

4. **ArrayExpress**
ArrayExpress – functional genomics data
http://www.ebi.ac.uk/arrayexpress/