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@6e36eca 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.

References