Local gene co-expression: a view from single-cell RNA-seq and ATAC-seq data

Diogo Ribeiro1,2, Chaymae Ziyani3, Olivier Delaneau1,2
1. Department of Computational Biology, University of Lausanne, Lausanne, Switzerland
2. Swiss Institute of Bioinformatics (SIB), Switzerland

Abstract

Nearby genes are often expressed as a group1. Yet, the prevalence, reasons and genetic control of local gene co-expression are far from being understood.

By leveraging gene expression measurements from the GTEx project across 49 human tissues and >800 individuals, we have previously found local gene co-expression to be highly prevalent2, occurring in 13% to 53% genes per tissue and comprising >64,000 co-expressed gene pairs (COPs).

To understand how local gene co-expression manifests at the single-cell level, here we analysed scRNA-seq across 87 genotyped individuals in a homogeneous cell type (iPSC)3,4. We found widespread local gene co-expression in single cells, particularly among functionally-related genes. Using bulk RNA-seq for the same 87 iPSC lines, we compared COPs identified with both technologies, revealing the potential of single cell analysis in retrieving unique gene co-expression patterns.

Finally, we explored the regulatory mechanisms involved in gene co-expression. Using scRNA-seq and scATAC-seq data from the same cells3, we identify enhancers associated with gene co-expression.

1. 1. scRNA-seq local gene co-expression (COP) detection
   1. Input data: scRNA-seq (total 7440 undifferentiated iPSC cells, Smart-seq2) from Cuomo et al 20214 and matching bulk RNA-seq for iPSC lines of 87 individuals (part of the HipSiC consortium)
   2. Calculated gene co-expression for all protein coding and lincRNA genes (250kb window, 5% FDR):
      - Bulk: a single matrix with individuals as columns, genes as rows
      - Per individual: a matrix per individual, cells as columns, genes as rows

3. Enhancers involved in gene co-expression
   1. Delaneau et al. (2019). Science, 364(6439)

Discussion & Conclusion

Besides having been found prevalent across tissues and cell lines, here we find COPs to be prevalent at the single cell level. Moreover, we found evidence for COP genes to be co-transcribed (GRO-seq data) and their co-expression to be kept up to the protein stage. The usage of a state-of-the-art dataset comprised of scRNA-seq & scATAC-seq in the same cells allowed us to identify enhancers associated with COPs and evidenced the role of regulatory element sharing in local gene co-expression. These and previous results indicate a link between trait pleiotropy of genetic variants due to shared gene regulation, yet, such a link remains unassessed.

Spearman correlation (p > 0.05, FDR 5%). Enhancer annotations specific to LCLs from EpiMap repository8

7. Mirauta et al. (2020) eLife, 10;9:e57390