

Diogo Ribeiro<sup>1,2</sup>, Chaymae Ziyani<sup>1</sup>, Olivier Delaneau<sup>1,2</sup>

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 group<sup>1</sup>. 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 prevalent<sup>2</sup>, occurring in 13% to 53% genes per tissue and comprising >64.000 co-expressed gene pairs (COPs).

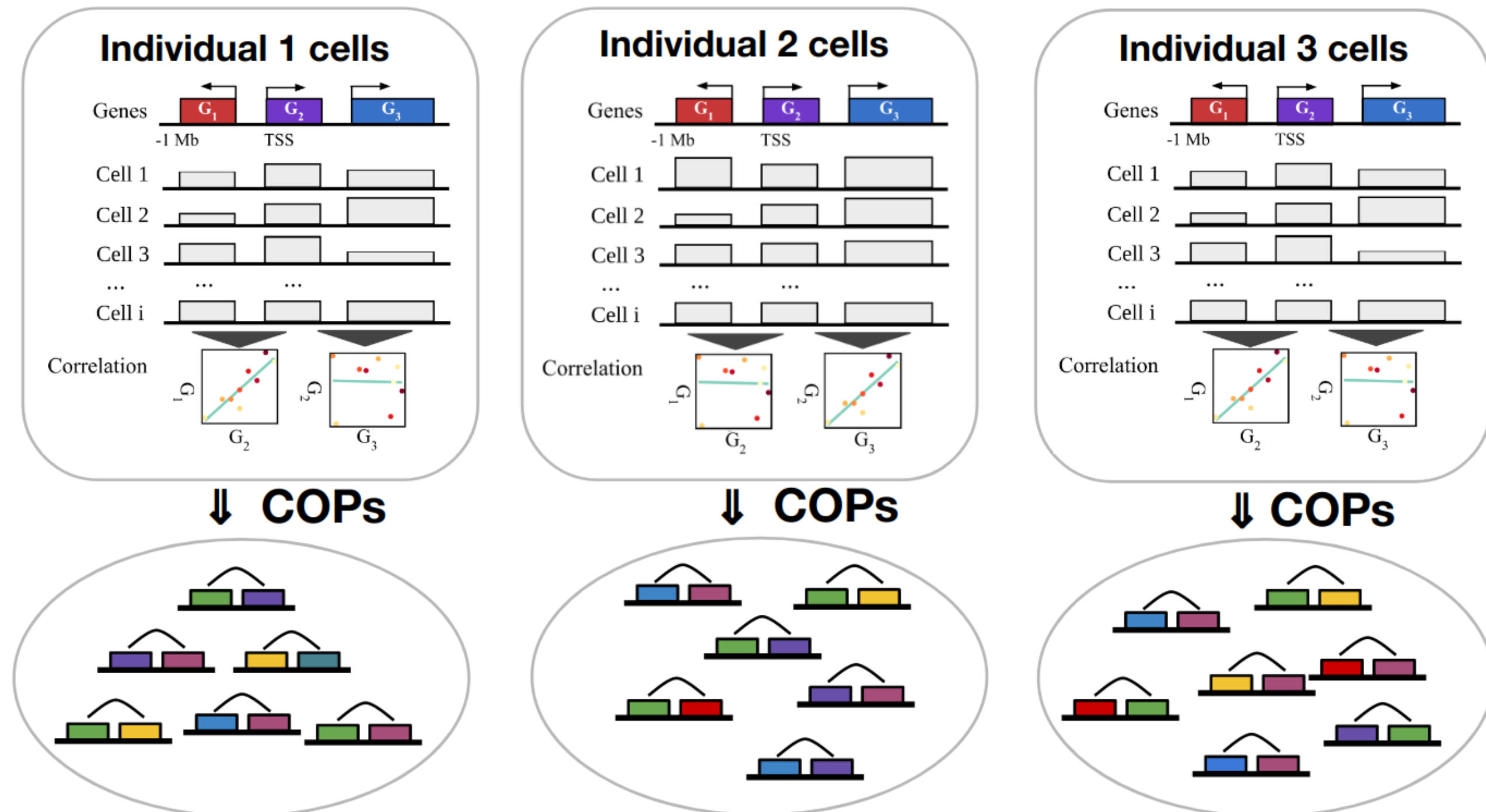
To understand how local gene co-expression manifests at the single-cell level, here we analysed scRNA-seq across 87 genotyped individuals in a homogeneous cell type (iPSC)<sup>3,4</sup>. We found widespread local gene co-expression in single cells, particularly among functionally-related genes. Using 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 cells<sup>5</sup>, we identify enhancers associated with gene co-expression.

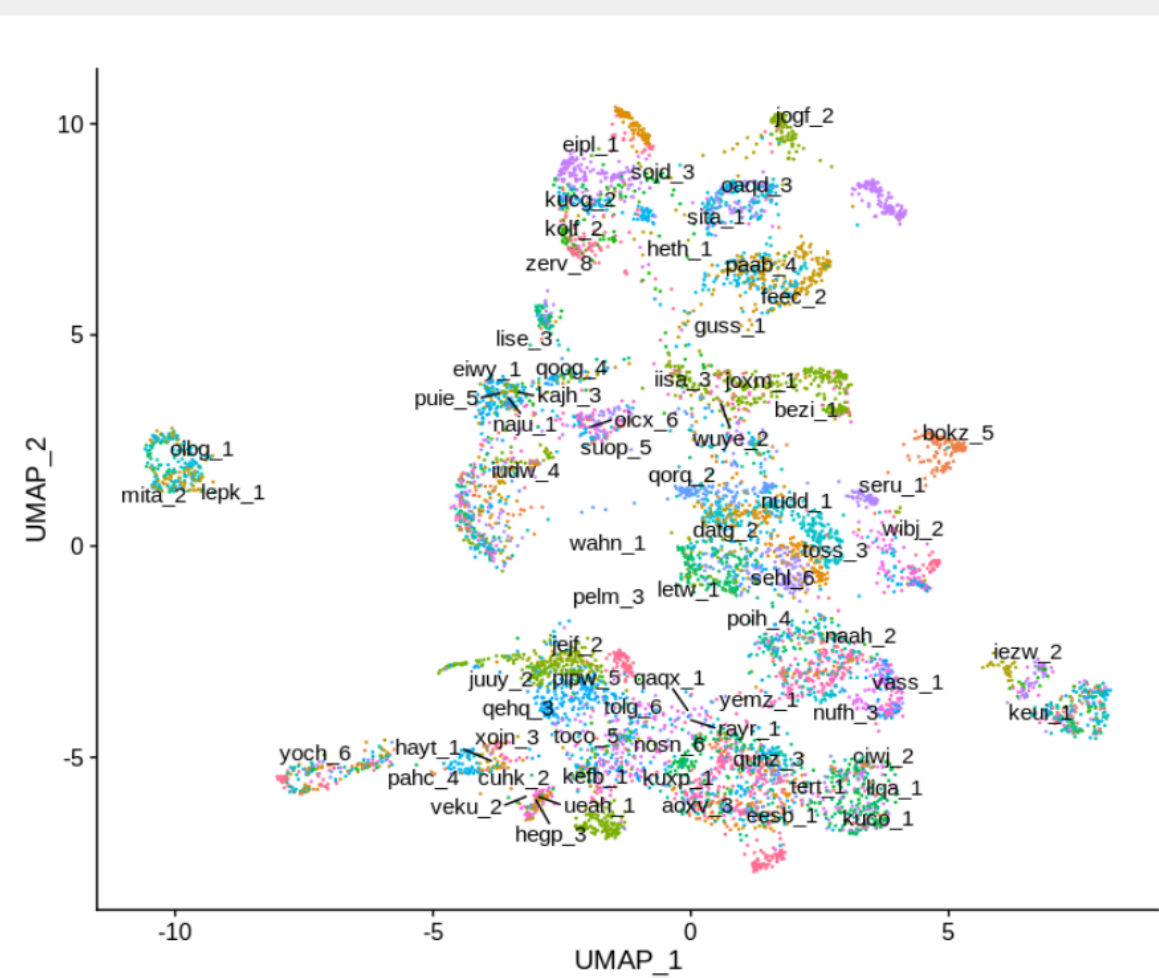
## 1. scRNA-seq local gene co-expression (COP) detection

- Input data: scRNA-seq (total 7440 undifferentiated iPSC cells, Smart-seq2) from Cuomo *et al* 2021<sup>4</sup> and matching bulk RNA-seq for iPSC lines of 87 individuals (part of the HipSci consortium)
- 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

### COP discovery per individual

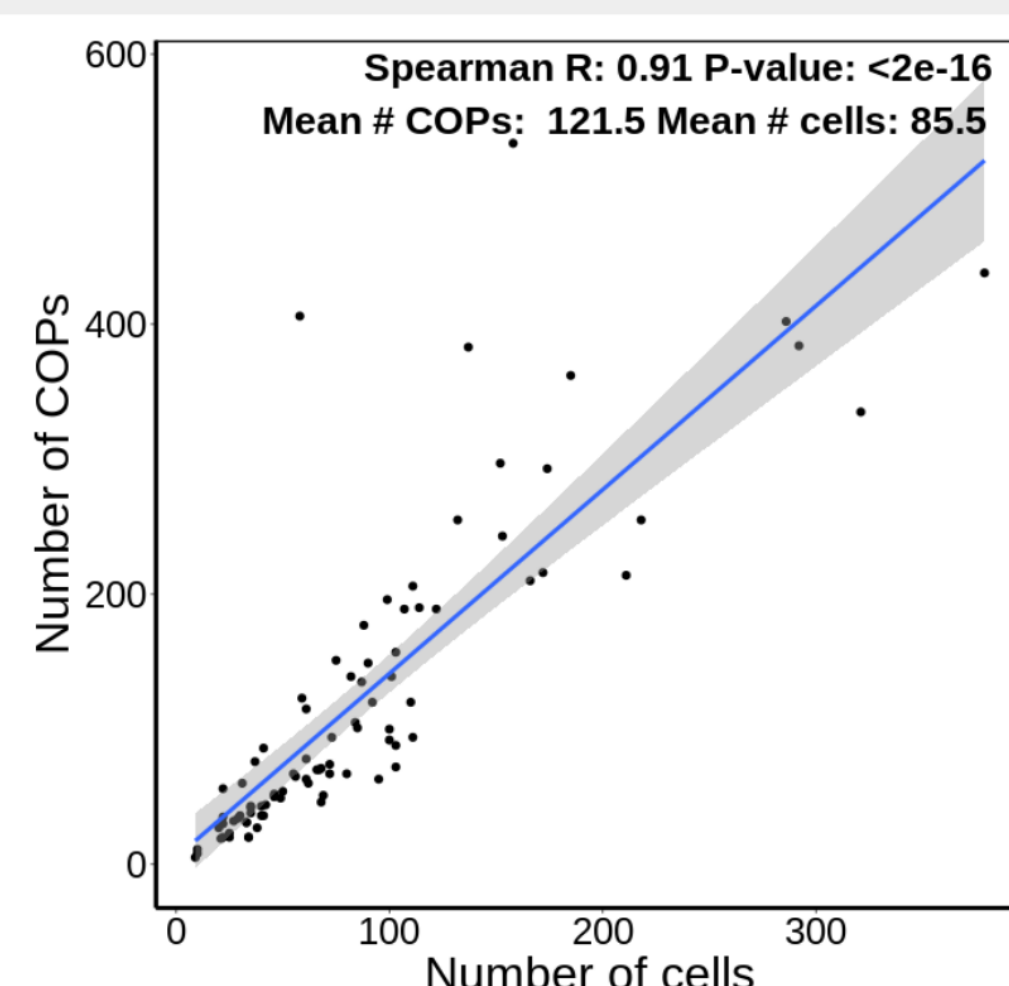


### Gene expression UMAP colored by individual



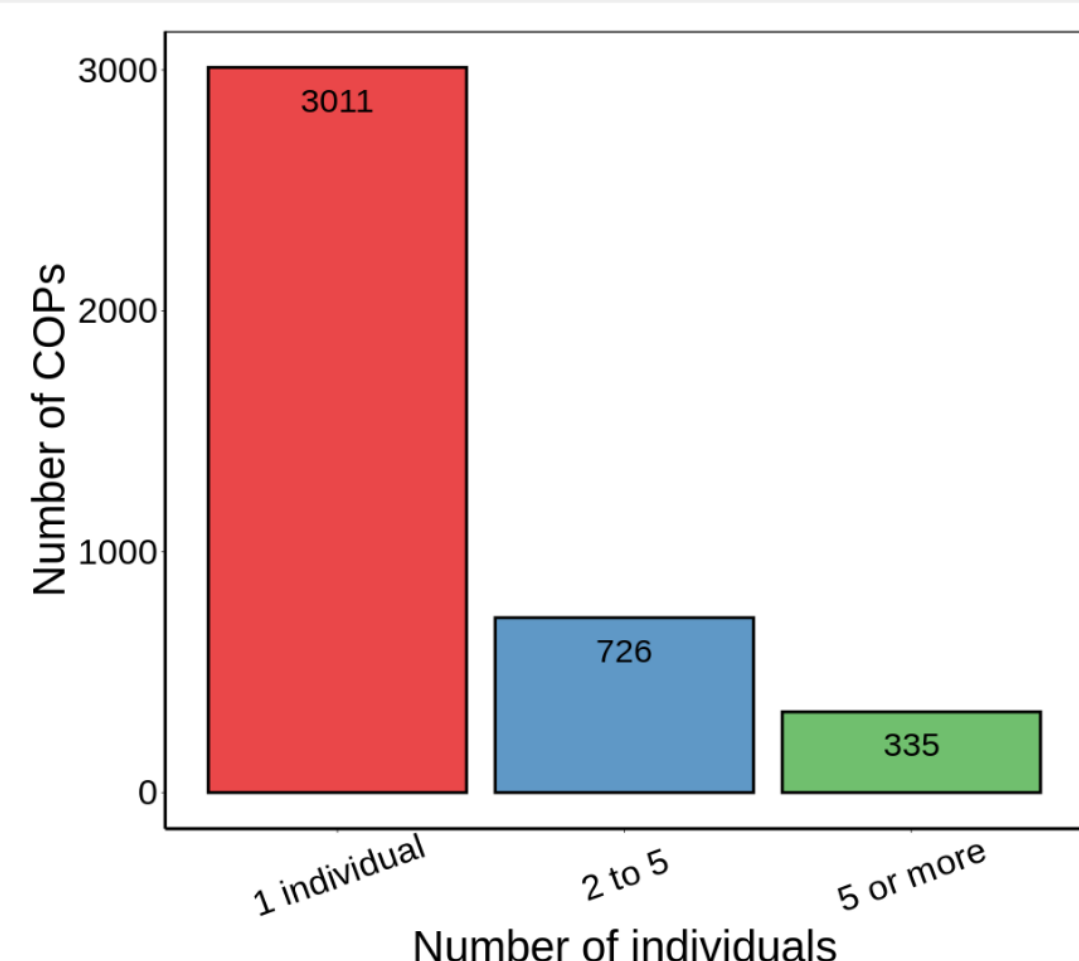
Cells are grouped per individual and per experimental batch (not shown) relatively well

### COP discovery across 87 individuals



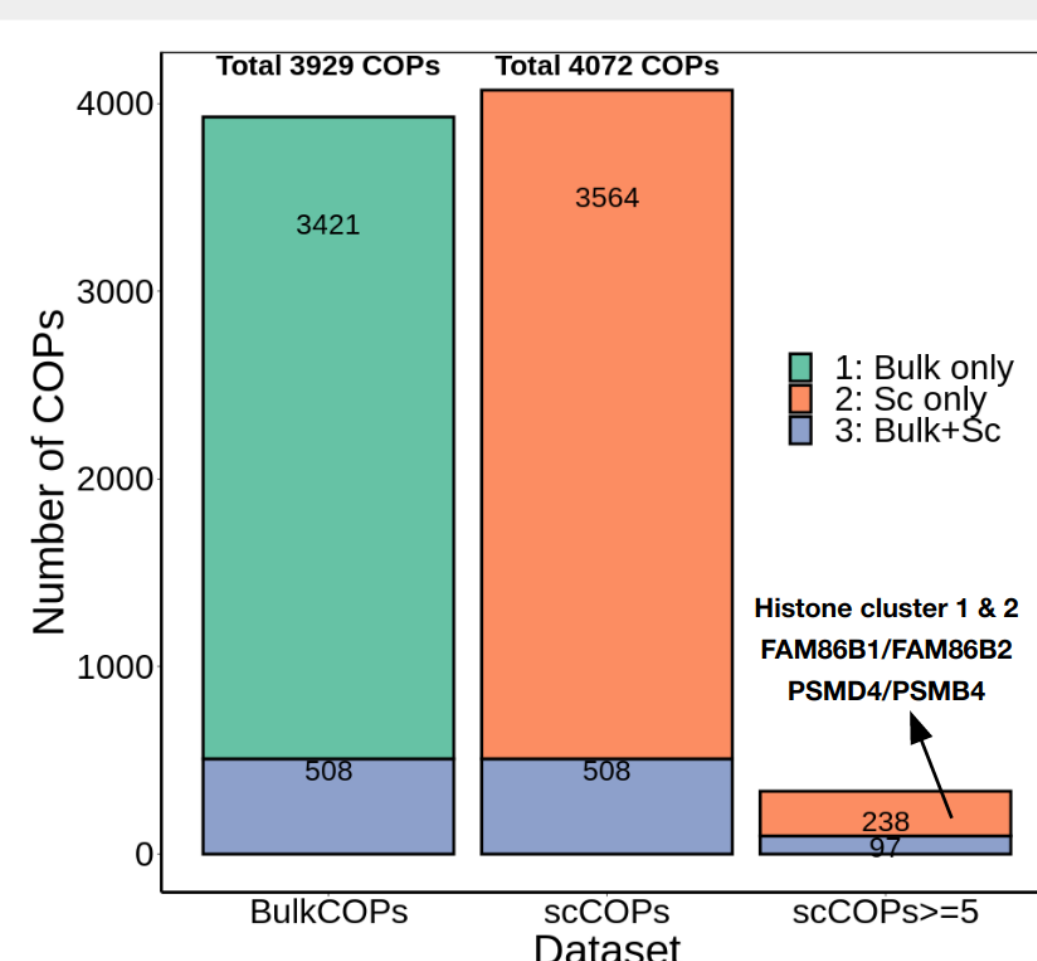
Number of COPs identified depends on the number of per cells per individual

### COP 'conservation' across individuals



Small fraction of conserved COPs: discovery power or high individual specificity?

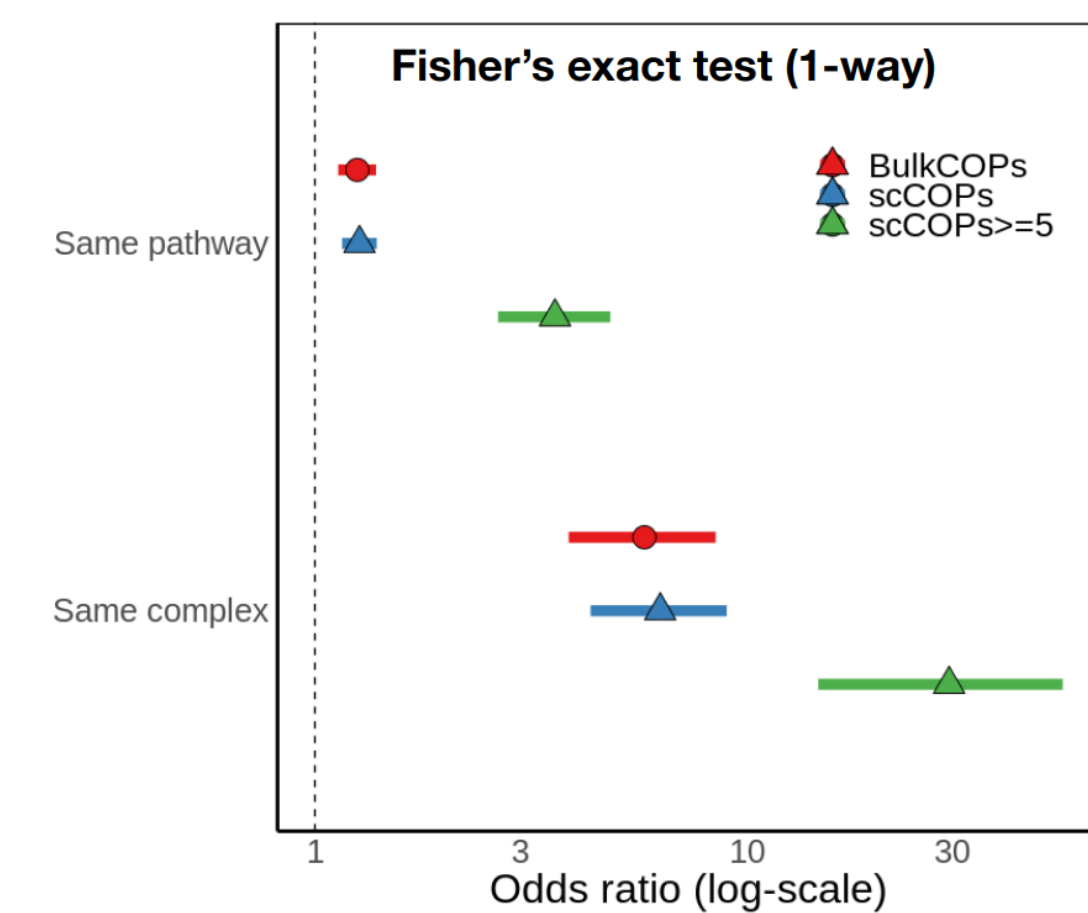
### Overlap between bulkCOPs and scCOPs



Similar total # COPs between bulk and single cell data, but low overlap between them

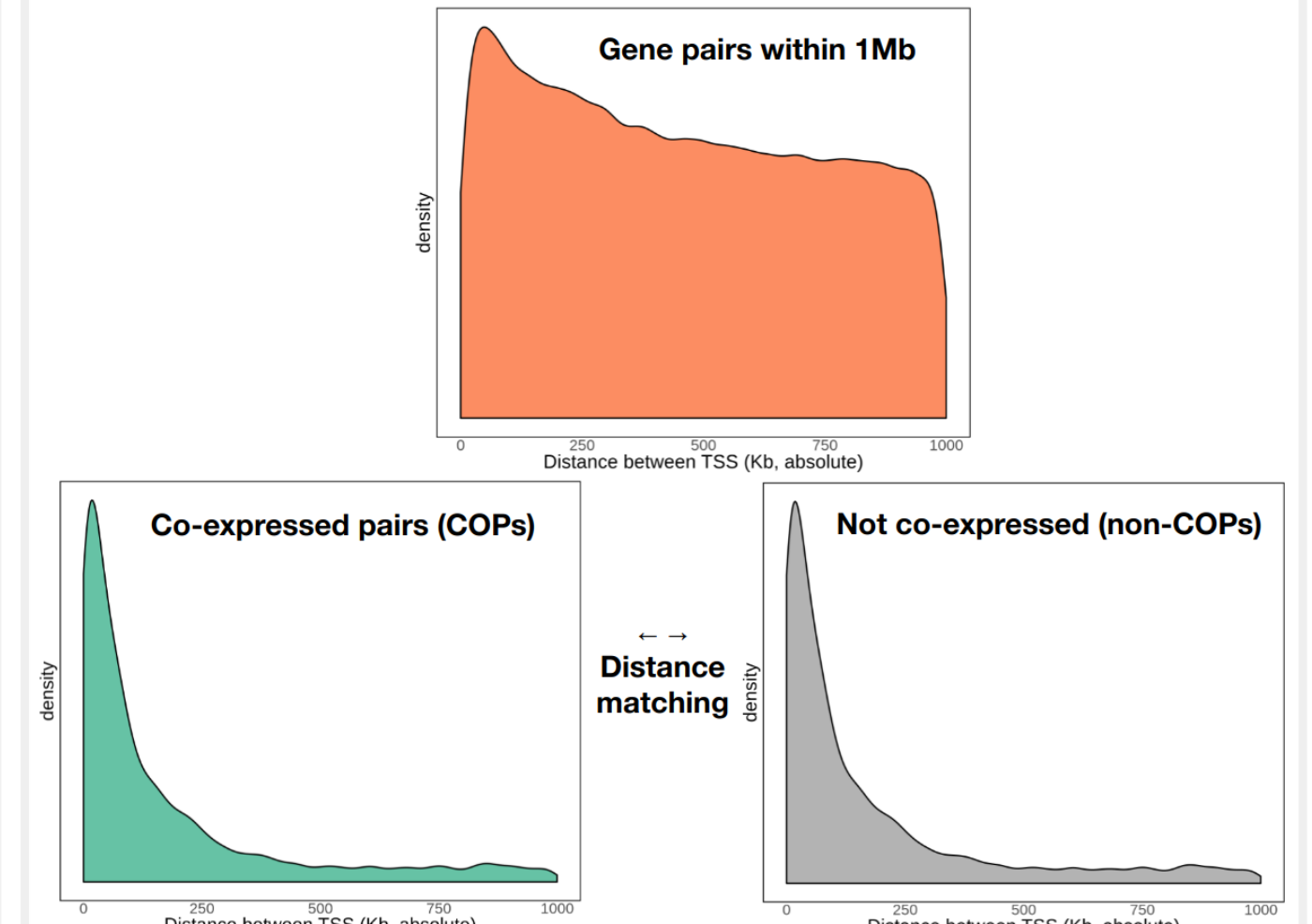
## 2. Properties of co-expressed gene pairs

### Enrichment of gene pairs in the same protein complex and same pathway



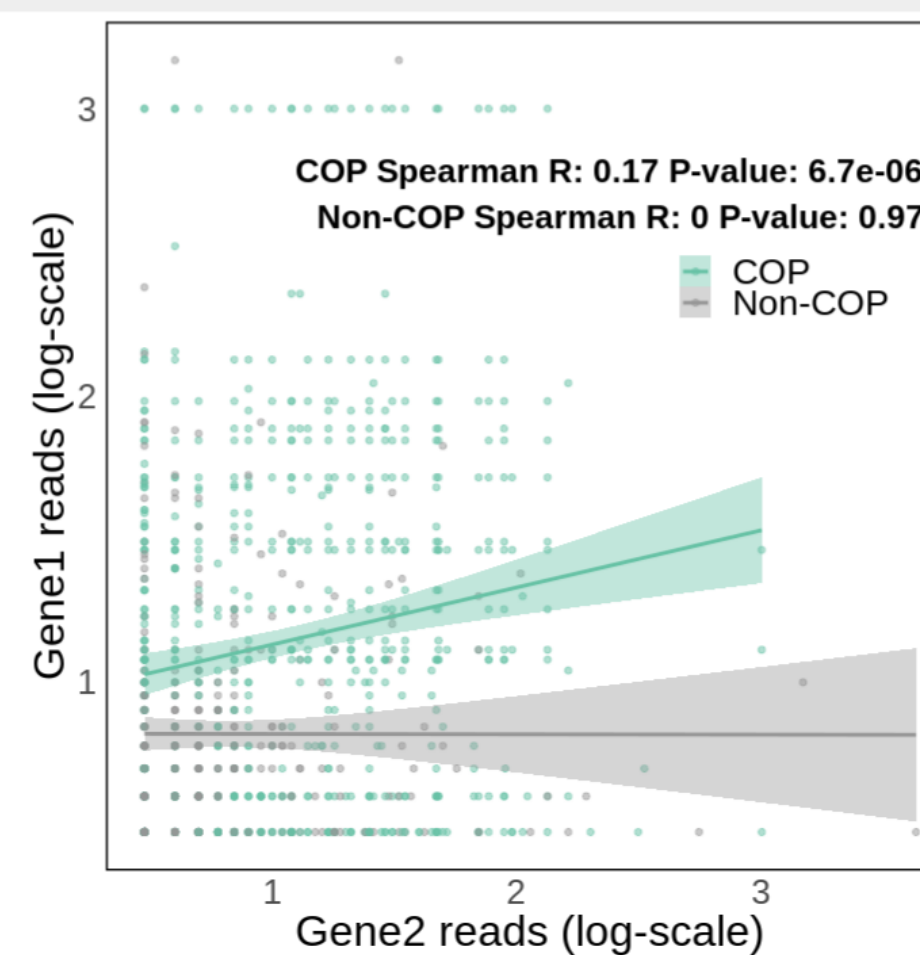
scCOPs are enriched in the same pathway (Reactome, KEGG) and protein complex (CORUM, Hu.Map)

### Distance matching between gene TSSs



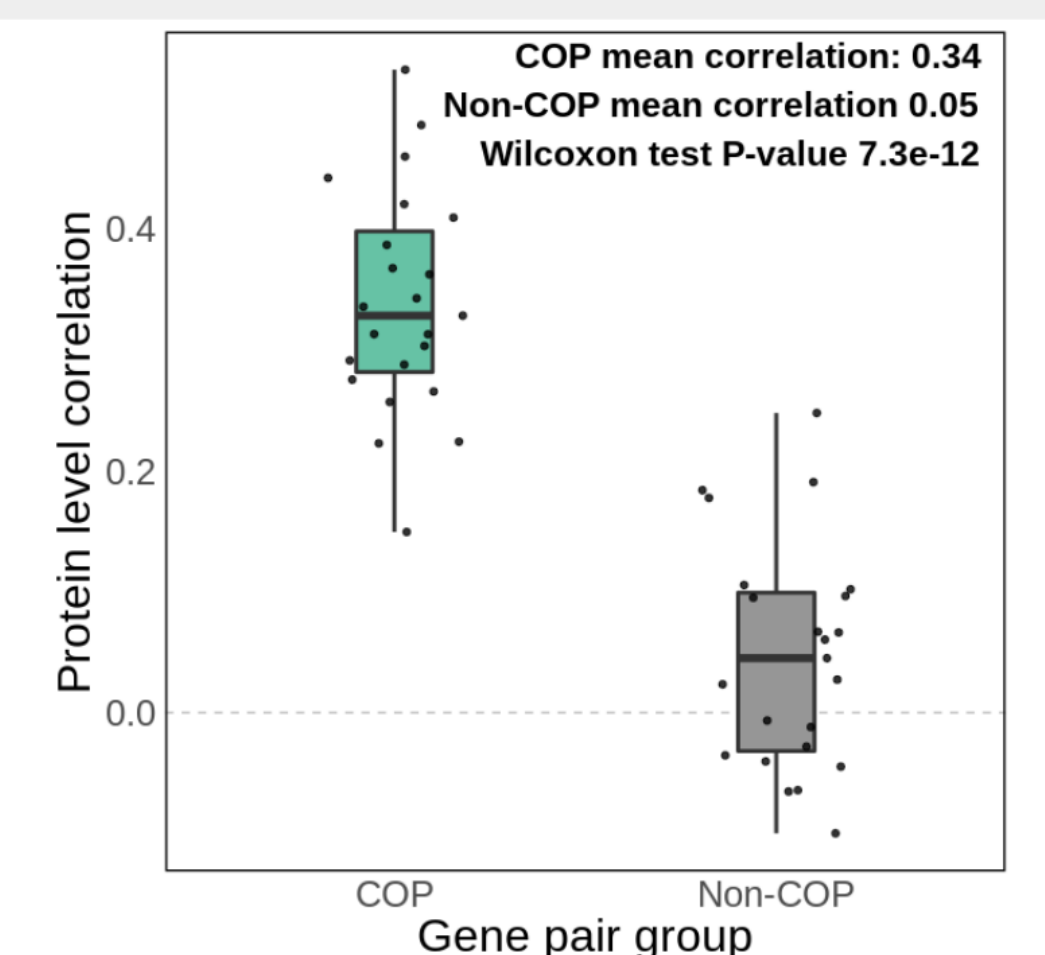
Most COPs are composed of nearby genes  
Control non-COPs match COP's distance distribution

### Correlation of gene pair transcription (GRO-seq data<sup>6</sup>)



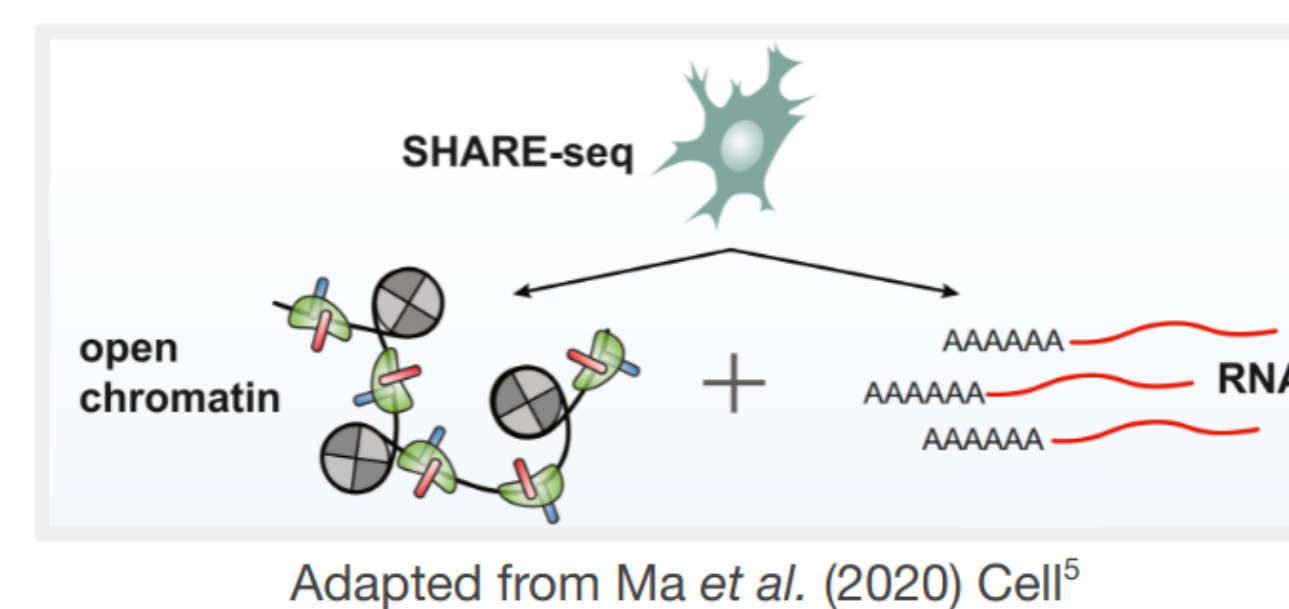
COPs found as nascent transcripts suggests simultaneous transcription

### Correlation of protein intensities (23 matching individuals, proteomics data<sup>7</sup>)



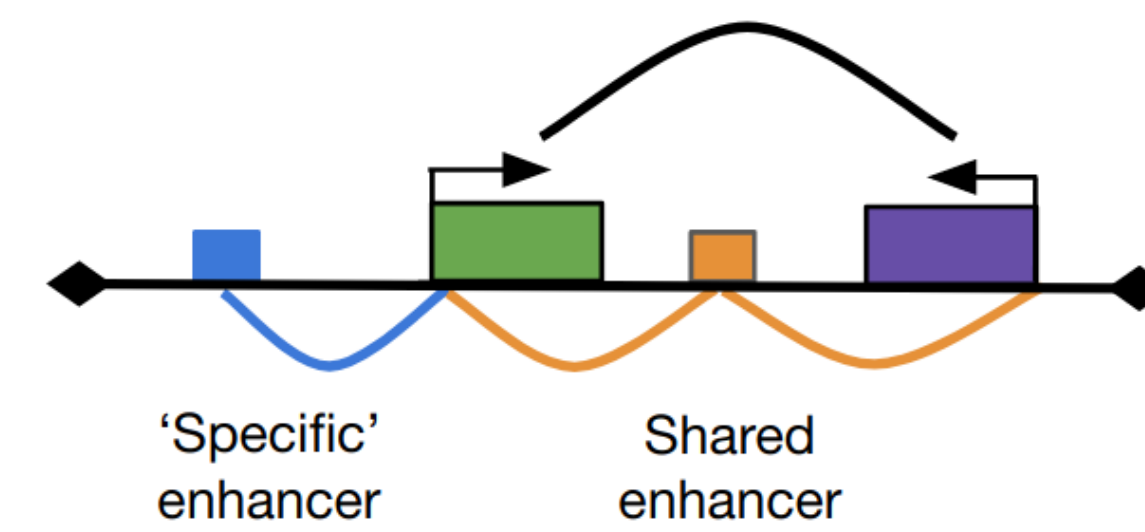
Proteomics data shows that COP gene co-expression is consistent at the protein level

## 3. Enhancers involved in gene co-expression

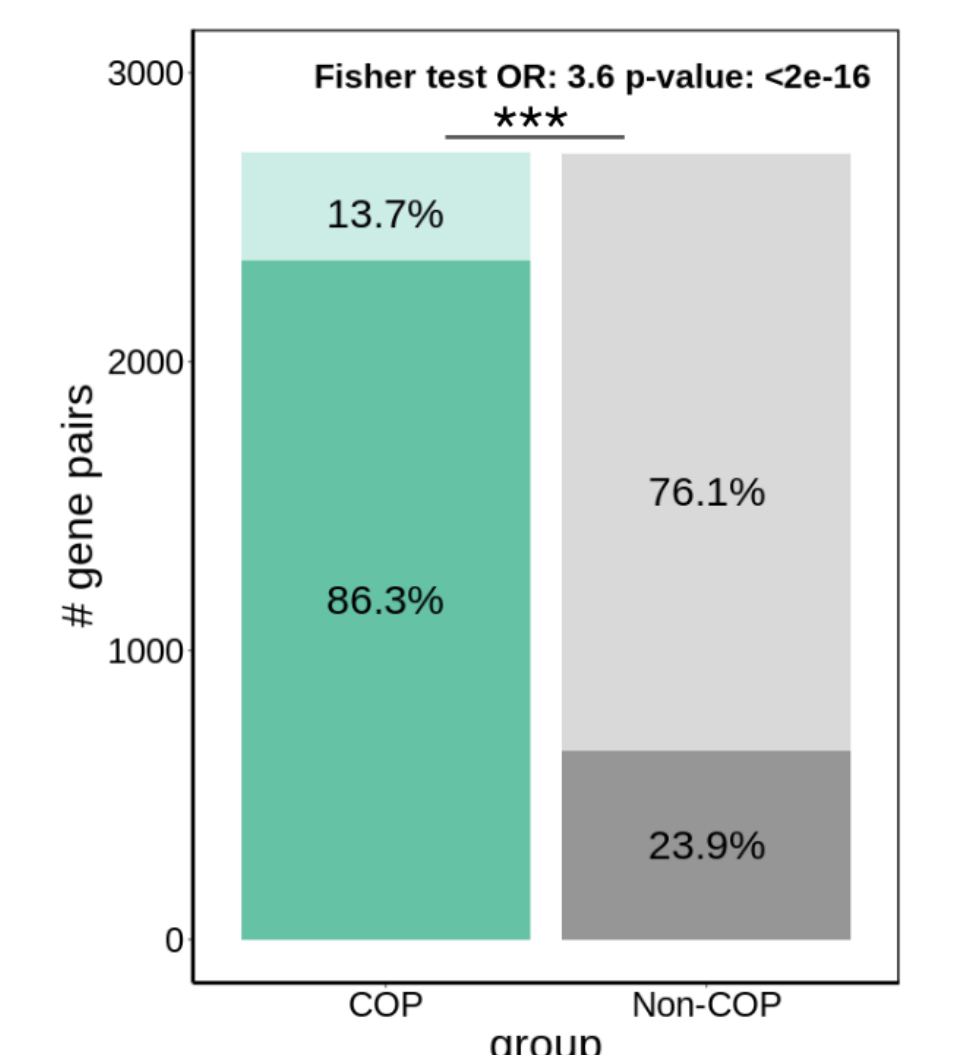


Adapted from Ma *et al.* (2020) Cell<sup>5</sup>

### Enhancer sharing



### Gene pairs with enhancer sharing



COPs display enhancer sharing more often than non-COPs. >85% COPs sharing enhancers

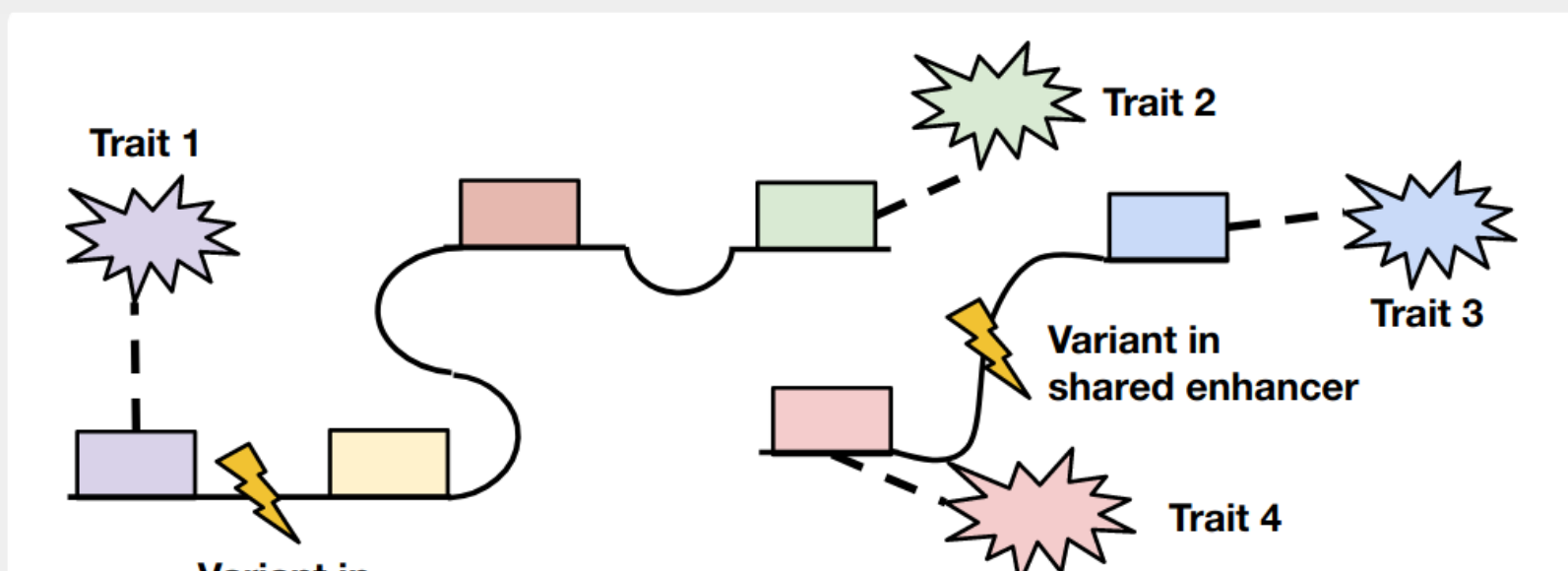
scRNA-seq and scATAC-seq in the very same set of 23278 LCL cells. Enhancer-gene associations: Spearman correlation ( $p > 0.05$ , FDR 5%). Enhancer annotations specific to LCLs from EpiMap repository<sup>8</sup>

## 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.

### Interpretation of eQTL/GWAS: one variant ≠ one gene



Could genetic perturbations affecting COPs (e.g. through enhancer sharing) lead to the disruption of multiple traits?