A single cell eQTL atlas for cell type specific regulatory effects

Lida Wang[1], Havell Markus[2,3], Dieyi Chen[1], Chachrit Khunsriraksakul[2,3], Laura Carrel[4], Dajiang J. Liu[1,2,4]

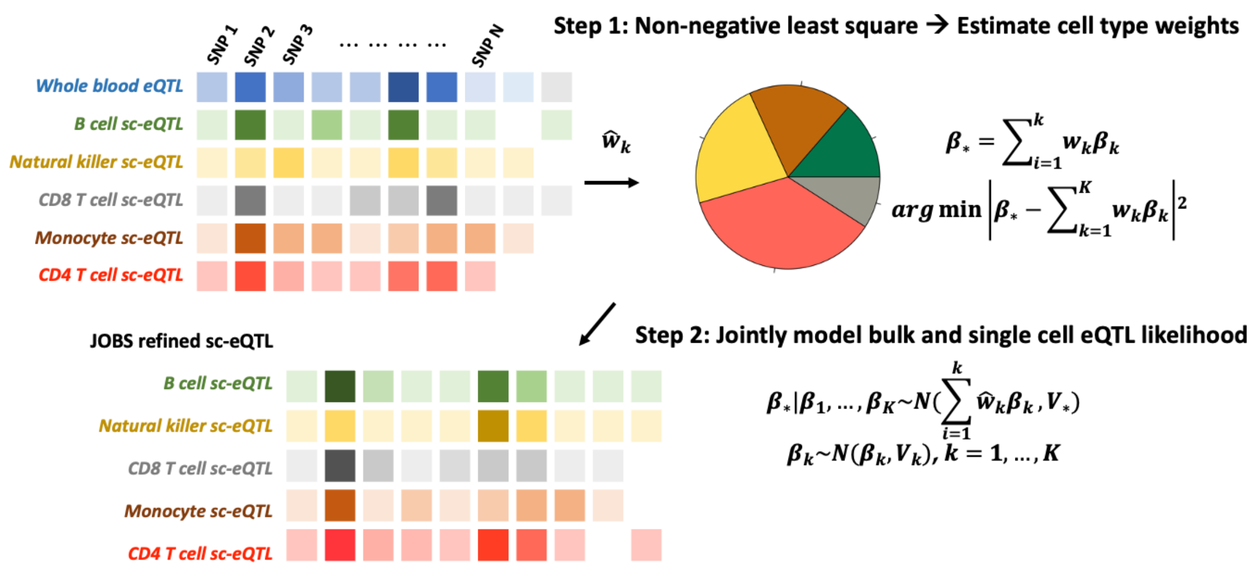
1 Department of Public Health Sciences; Pennsylvania State University College of Medicine; Hershey, Pennsylvania, 17033; USA.
2 Bioinformatics and Genomics PhD Program; Pennsylvania State University College of Medicine; Hershey, Pennsylvania, 17033; USA.
3 Institute for Personalized Medicine; Pennsylvania State University College of Medicine; Hershey, Pennsylvania, 17033; USA.
4 Department of Biochemistry and Molecular Biology; Pennsylvania State University College of Medicine; Hershey, Pennsylvania, 17033; USA

Background

GWAS has discovered numerous associations with complex traits and diseases. Most associated variants are non-coding and function by regulating gene expressions. Dissecting mechanisms of the GWAS loci requires integrating eQTL data using colocalization analysis or TWAS. To date, many GWAS hits still fail to colocalize with eQTL variants, possibly due to the lack of power in eQTL or GWAS studies and the regulatory effect differences between cell types which are not captured by bulk RNASeq data.

Single-cell RNASeq (scRNASeq) can systematically characterize cell type specific regulatory effects of gene expression. By grouping cells into pseudo-bulks of different cell types, cell type specific eQTLs (sc-eQTLs) can be calculated. The success of TWAS and colocalization critically depends on the power of sc-eQTL studies. The largest sc-eQTL datasets contain 1000s of individuals, which are still much smaller than bulk RNASeq data. To improve the power of sc-eQTL analysis, we develop a new method **JO**int analysis of **B**ulk and **S**c-eQTL (**JOBS**). JOBS relies on the key insight that bulk eQTL datasets can be viewed as a weighted average of sc-eQTLs from constituting cell types. Jointly analyzing bulk- and sc-eQTL datasets with a compositional model will borrow strength from the large sample size of bulk eQTL datasets and substantially improve sc-eQTL effect estimates. JOBS generates more accurate sc-eQTL effect estimates, which can in turn improve all downstream analysis, including eQTL discovery, TWAS, and co-localization analysis.

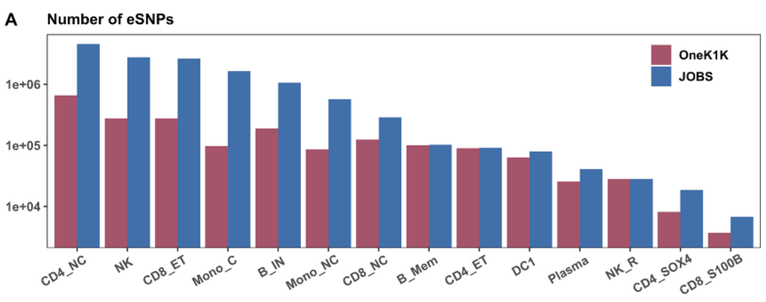
Method: JOint analysis of Bulk and Sc-eQTL (JOBS)



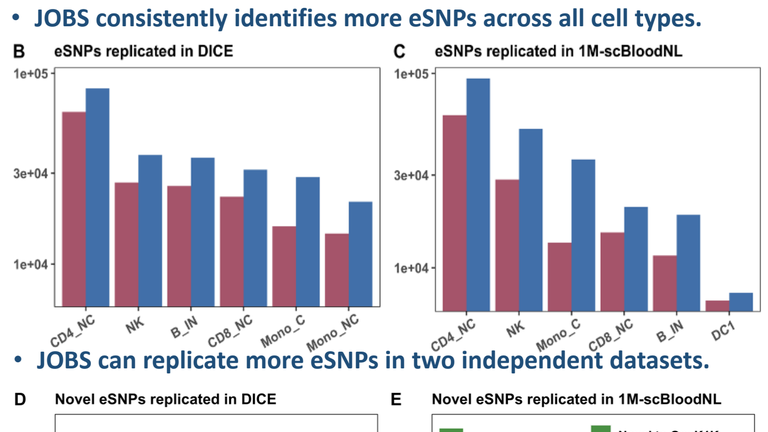
Results

Data:

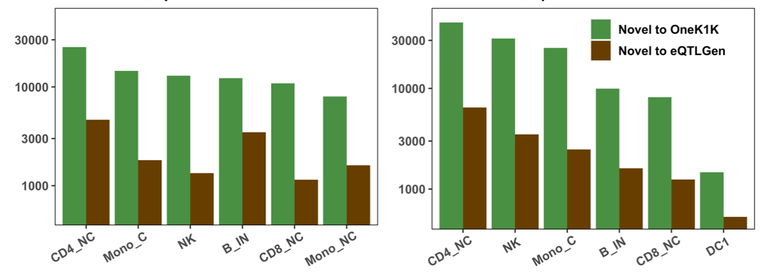
1. **OneK1K data** - a large sc-eQTL data of 14 cell types with N =982.
2. **eQTLGen data** - a large bulk eQTL data with N = 31684).



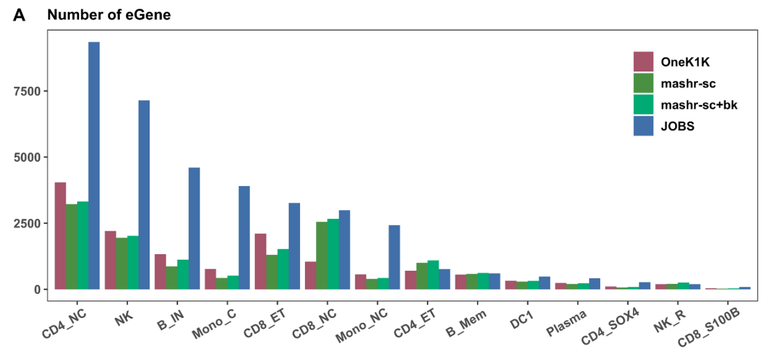
- **JOBS consistently identifies more eSNPs across all cell types.**



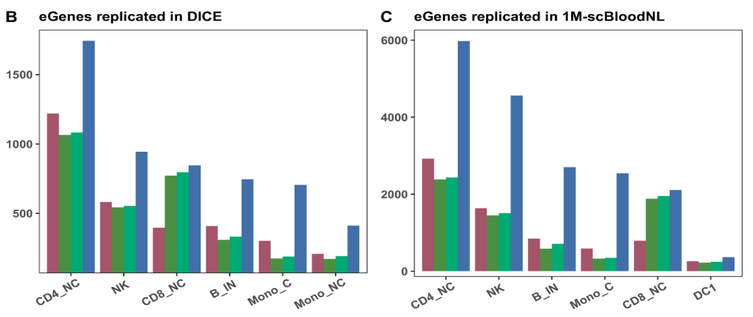
- **JOBS can replicate more eSNPs in two independent datasets.**



- **JOBS can also replicate more novel eSNPs.**



- **JOBS identifies more sc-eGenes than *mashr*.**



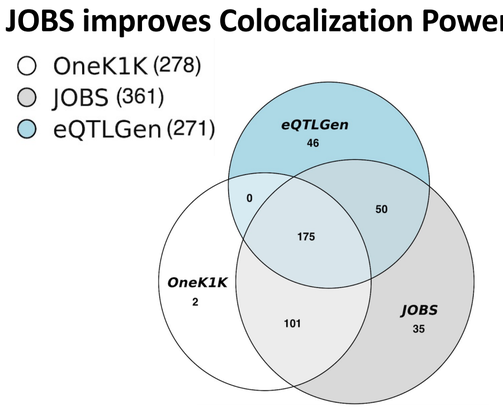
- **JOBS also replicates more sc-eGenes than *mashr*.**

- **We can represent the improved power by an equivalent improvement in the sample sizes**

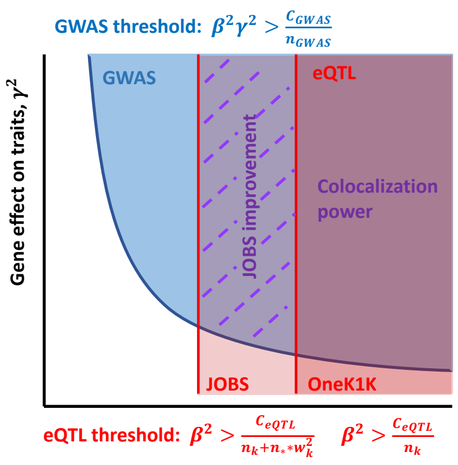
$$N_{JOBS} = N_{OneK1K} + w_i^2 N_{eQTLGen}$$

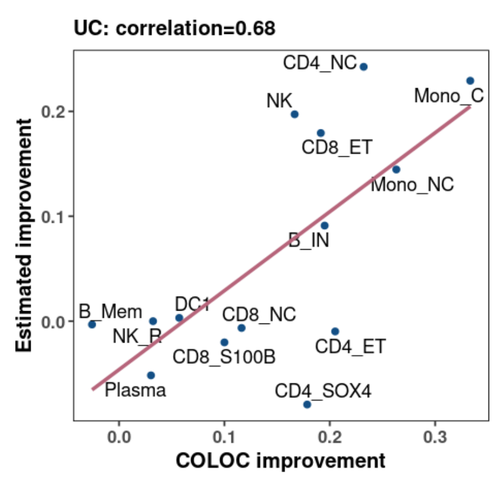
- **JOBS offers bigger improvement for more common cell types.**

JOBS improves Colocalization Power



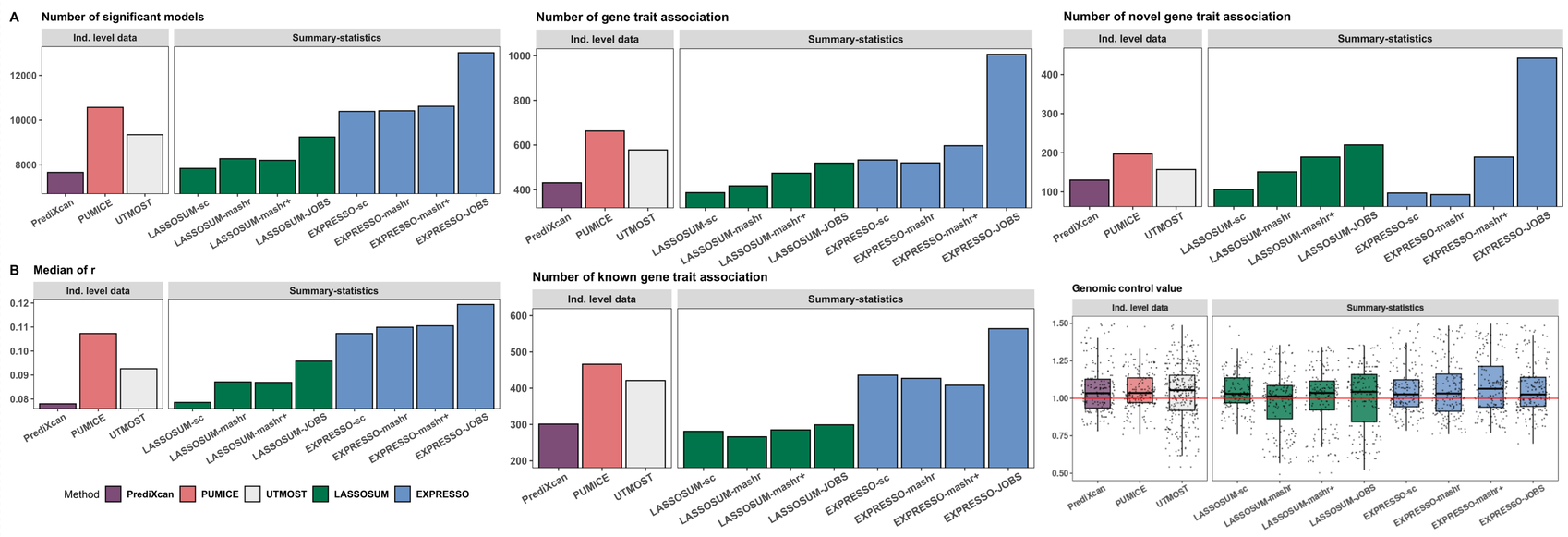
- **361 unique loci, which is 29.86% more than sc-eQTLs in OneK1K and 33.21% more than the number of eQTLs from eQTLGen.**





- **The observed improvement in the number of colocalized loci is consistent with the theoretical expectation based on improved sc-eQTL analysis power.**

JOBS improves in-silico gene expression prediction and TWAS power

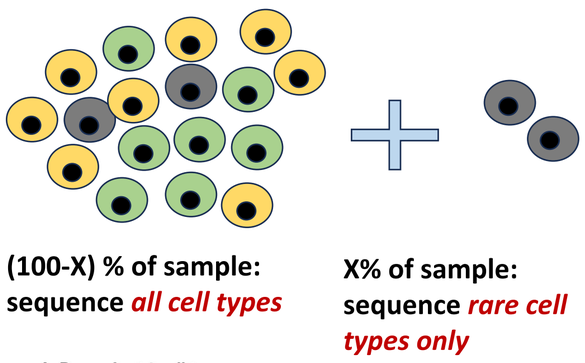


- **More significant prediction models.**
- **improved gene expression prediction accuracy.**
- **Identify more risk genes.**
- **Replicate more genes in GWAS catalog.**
- **Well controlled type I errors.**
- **Identify more novel genes.**

Optimal sc-RNAseq design to maximize the power of sc-eQTL mapping

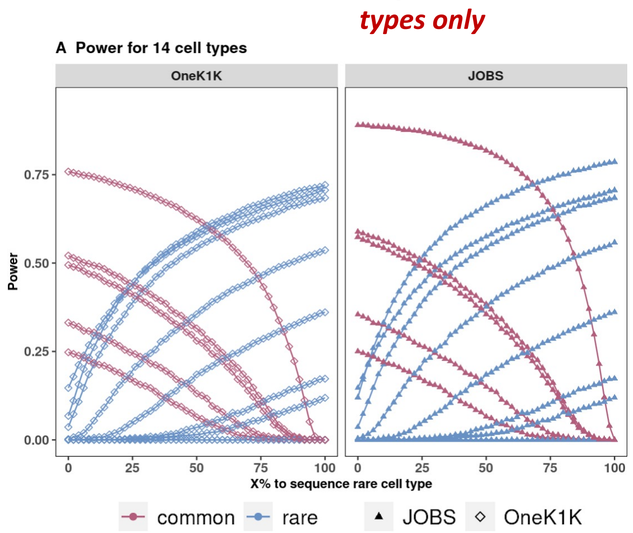
- **JOBS substantially improves sc-eQTL mapping for common cell types.**
- **Yet, the improvements for rare cell types are smaller.**
- **Increasing sequence depths for rare cell types improves sc-eQTL mapping power.**

Novel sc-RNAseq strategy to maximize sc-eQTL power



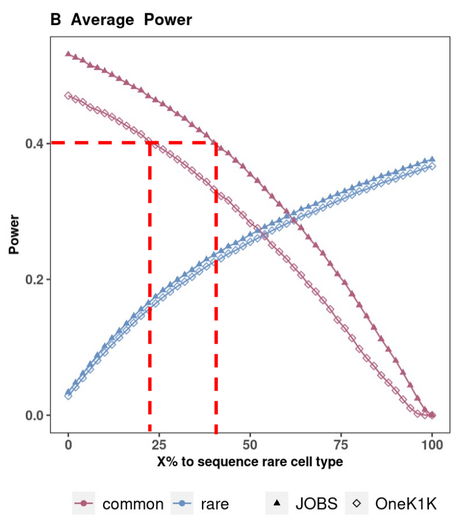
(100-X) % of sample: sequence *all cell types*

X% of sample: sequence *rare cell types only*



As X increase, the power for rare cell types increases

Question: to ensure Y% of power of sc-eQTL mapping for **common** cell types, what is the power we can attain for **rare** cell types?



Applying JOBS allows us to sequence a smaller fraction of samples for all cell types →
We can sequence more samples for rare cell types→
Improves the power by 43% for rare cell types.

Methods (X%)	Common	Rare
1k1k (20%)	0.4	0.165
JOBS(40%)	0.4	0.236

Conclusion

- **Developed novel method: JOBS (JOint analysis of Bulk and Sc-eQTLs).**
- **JOBS could benefit eQTL discovery and downstream analysis, including colocalization and TWAS.**
- **Proposed an optimal sc-RNAseq design to maximize the power of sc-eQTL mapping.**