

# DSS Statistics Seminar

May 7, 2021, 15:00

<https://uniroma1.zoom.us/j/86881977368?pwd=SWRFcVFjMDZTa0lXZk05TE1zNm5adz09>

Passcode: 432940

Identifying condition-specific patterns in large-scale genomic data

## *Qunhua Li*

*Dept of Statistics, Penn State University*

Joint analyses of genomic datasets obtained in multiple different conditions are essential for understanding the biological mechanism that drives tissue-specificity and cell differentiation. But it still remains computationally challenging even when the number of conditions is moderate.

I will present CLIMB (Composite Likelihood eMpirical Bayes), a statistical methodology which learns patterns of condition specificity present in genomic data by leveraging pairwise information. CLIMB provides a generic framework facilitating a host of downstream analyses, such as clustering genomic features sharing similar conditional-specific patterns and identifying which of these features are involved in cell fate commitment. It improves upon existing methods by boosting statistical power to identify biologically meaningful signals while retaining interpretability and computational tractability. We illustrate CLIMB's value on a CTCF ChIP-seq dataset measured in 17 different cell populations and an RNA-seq dataset measured in three committed hematopoietic lineages. These analyses demonstrate that CLIMB captures biologically relevant clusters in the data and improves upon commonly-used pairwise comparisons and unsupervised clusterings typical of genomic analyses.



**SAPIENZA**  
UNIVERSITÀ DI ROMA