



THE UNIVERSITY OF  
CHICAGO

DEPARTMENT OF STATISTICS

## Statistics Colloquium

---

ZONGMING MA

Department of Statistics and Data Science  
University of Pennsylvania

“Matching and integration of datasets with low-rank signals and applications in single-cell data analysis”

MONDAY, OCTOBER 24, 2022, at 4:30 PM

Jones 303, 5747 S. Ellis Avenue

*Refreshments before the seminar at 4:00 PM in Jones 304.*

### ABSTRACT

We study one-way matching of a pair of datasets with low rank signals. Under a stylized model, we first derive information-theoretic limits of matching under a mismatch proportion loss. We then show that linear assignment with projected data achieves fast rates of convergence and sometimes even rate optimality for this task. Built upon this theoretical understanding, we propose a new matching algorithm, Matching with pARTial Overlap (MARIO), that takes into account both shared and distinct features in related datasets. MARIO accurately matches and integrates data from different single-cell methods. As an exemplary application, it robustly matched tissue macrophages identified from COVID-19 lung autopsies via CODEX imaging to macrophages recovered from COVID-19 bronchoalveolar lavage fluid via CITE-seq to provide evidence for orchestrated immune responses within the lung microenvironment of COVID patients.