# Improving Single-Cell Perturbation Analyses through Efficiency Estimation

Qiyuan Liu[1], Siming Zhao[2], **Jingshu Wang[1]**

[1] *Department of Statistics, The University of Chicago*

[2] *Department of biomedical data science and the Cancer Center, Dartmouth College*

## ABSTRACT

Single-cell perturbation screening has transformed functional genomics, yet its utility is hampered by noisy data and low statistical power. A central but often overlooked challenge is perturbation efficiency: some perturbations strongly alter target gene expression, while others fail to induce measurable changes, even within the same experiment. Ignoring this heterogeneity renders causal estimands dataset-dependent and undermines reproducibility.

We show that existing approaches for estimating cell-level perturbation efficiency are vulnerable to confounding, leading to biased effect estimates and false discoveries. To address this, we develop an instrumental-variable framework that treats perturbations as instruments rather than direct treatments. By estimating gRNA-specific efficiencies—shifting resolution from individual cells to groups of cells—we obtain consistent estimates that reveal striking variability across gRNAs targeting the same gene. Incorporating these efficiencies into differential expression analyses improves detection power, yields more interpretable causal effect sizes, and enhances consistency across datasets. When sample sizes permit, our framework further enables investigation of dosage effects, providing a clearer picture of heterogeneity in gene perturbation responses.

Our results highlight the importance of rigorously defining and estimating perturbation efficiency, thereby improving both the validity and interpretability of single-cell perturbation studies.

**Keywords:** instrumental variable, single-cell genomics, CRISPR screening, causal effect estimation

# Addressing heterogeneous sensitivity in biomarker screening with application in NanoString nCounter data

Chang Yu, **Zhijin Wu**

*Department of Biostatistics, Brown University*

**ABSTRACT**

Biomarkers are measurable indicators of biological processes and have wide biomedical applications including disease screening and prognosis prediction. Candidate biomarkers can be screened in high-throughput settings, which allow simultaneous measurements of a large number of molecules. The ability to detect a molecule may be hindered by the presence of background noise and the variable signal strength, which depends on both the properties of the target molecule and the quality of the sample. The detection sensitivity thus varies in a marker- and sample-specific manner. This heterogeneity in detection sensitivity is often overlooked and leads to an inflated false positive rate. We propose a novel *sensitivity adjusted likelihood-ratio test* (SALT), which properly controls the false positives and is more powerful than the unadjusted approach. We show that sample-and-feature-specific detection sensitivity can be well estimated from NanoString nCounter data, and using the estimated sensitivity in SALT results in improved biomarker screening.

Keywords: High-throughput screening, Biomarker, Sensitivity, NanoString nCounter

# The Taiwan Precision Medicine Initiative: Building the Largest non-European Cohort for Precision Health

**Jer-Yuarn Wu**

*Institute of Biomedical Sciences, Academia Sinica*

**ABSTRACT**

Data from the national biobanks, especially those from the UK Biobank, show that disease risk prediction based on genetic profiles is feasible but the prediction models created with European data do not translate to other ethnic groups. The current bottleneck is therefore a large dataset which consists of genetic profiles and matching clinical information from many individuals in Taiwan. Equally important is the analysis of this large dataset to produce disease risk models that predict an individual's disease risk for all the major diseases. To establish this large dataset, we collaborated with 16 major medical systems across Taiwan and launched Taiwan Precision Medicine Initiative (TPMI) in 2019.

We produced a reference panel against which anyone of Han Chinese ancestry can use to assess his/her own disease risk for precision health management. Leveraging the whole genome sequencing data produced by the Taiwan Biobank and the genome reference assemblies produced by the Kwok's group, we have designed and validated custom single nucleotide polymorphism (SNP) arrays that can produce the genetic profile of a person and test all known clinically useful genetic variants at the same time. We have genotyped ~500,000 participants enrolled (as of December 31, 2023). All participants have agreed to contribute their clinical data (EMR, electronic medical record) to the database, allowing us to analyze not just their current disease state, but also their treatment response and long term disease progression.

Our result showed that 327 diseases have more than 10,000 participants. 37 of those diseases have polygenic risk scores with high accuracy (AUC value larger than 0.6). We anticipate the established TPMI dataset will advance the promise of "Precision Medicine" for individuals in Taiwan.

**Key words:** polygenic risk score, EMR (electronic medical record), AUC

# A Multi-Tissue Map of Protein Regulation Reveals Shared and Context-Dependent Genetic Architectures

**Hua Tang**[1], Huaying Fang[1,2], Lihua Jiang[1]

[1]*Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA.*

[2]*Academy for Multidisciplinary Studies, Capital Normal University, Beijing 100048, China.*

## ABSTRACT

The abundance of proteins, essential functional units of the cell, is tightly controlled by genetic and environmental factors. Dysregulation of proteins is fundamental to disease. While transcriptomic and plasma proteomic studies have provided key insights into molecular regulation, a tissue-resolved regulatory map across normal human organs remains elusive. Here we quantify over 10,000 unique proteins across five disease-relevant tissues from hundreds of donors, establishing a systematic tissue proteome map. We identify nearly 2,000 cis-regulatory loci, most not previously linked to protein abundance in plasma, and show that these local effects are broadly shared across tissues. In contrast, trans-acting loci and sex associations are highly context-specific, while age associations show intermediate sharing. Integration with transcriptomics further reveals that, within a tissue, gene-level RNA–protein correlations across individuals are generally low, underscoring the pervasive role of post-transcriptional regulation. Together with existing plasma studies, these findings define a layered architecture of protein regulation that spans tissues and plasma, illuminating both shared and tissue-specific biology. This resource provides a molecular framework for connecting genetic and biological factors to protein regulation and for advancing mechanistic insight into human complex diseases.

**Keywords:** proteomics, multi-omics, normal human tissues, pQTL, gene-regulation