# A169 · Evaluating Computational Deconvolution Methods and Optimizing Gene Signature Matrices for Rare Cell Detection in Pediatric Cancer Liquid Biopsies

**Presenter:** Kenia Viri — Salve Regina University

**Authors:** Kenia Viri, Uma K. Paithankar, Roma Parika, Sahbra Eldosougi, Uyen K. Ho, Caroline Zielinski, Sarah San Vicente, Jacy Fang, J.M. Shim, Timothy West, Jenna Barboza, Mason Perry, Sofia Neary, Isaac Rivera, Jackson Brenek, Jacqueline Tremblay, Lecia Sequist, Brian Nahed, Moshe Sade-Feldman, Shannon L. Stott, Jillian F. Wise

**Session:** Day 1 · Thu Oct 1, 2026 · 2:15–4:15 PM

Liquid biopsy is a minimally invasive approach for monitoring disease through rare circulating cells in blood. In pediatric brain cancer, detecting low frequency populations, including circulating tumor cells (CTCs), is challenging due to their scarcity and the complex cellular composition of blood. While single-cell RNA sequencing (scRNA-seq) provides high resolution characterization, its cost and time requirements make computational deconvolution an important approach for estimating cell type proportions. Our lab is developing a pipeline for pediatric liquid biopsy analysis by integrating optimized deconvolution with rare cell detection. To evaluate existing approaches, we generated pseudobulk datasets from scRNA-seq data from ten healthy pediatric blood samples. Each dataset contained varying proportions of twenty annotated cell types and sequential spike-ins of 0-100 vascular endothelial cells to simulate rare populations. We evaluated CIBERSORT, CIBERSORTx, Decon, ReDeconv, and DWLS using Pearson's correlation, root mean square error (RSME), and mean absolute deviation (mAD). CIBERSORTx Fractions demonstrated the strongest performance for predicting cells present at or below 0.03%, achieving a correlation of 0.52, while ReDeconv showed decreased correlation as rare cell proportions increased. All methods struggled to predict cell populations at 0%. To further optimize deconvolution, gene signature matrices were generated from healthy pediatric blood scRNA-seq datasets. Gene selection and normalization were evaluated using heatmaps and linear discriminant analysis (LDA), with log-normalized data achieving the highest accuracy compared with raw counts and variance stabilizing transformation (VST). These findings establish performance evaluations and optimize reference profiles for developing SparCell to improve rare CTC detection in pediatric biopsies.

# Evaluating Computational Deconvolution Methods and Optimizing Gene Signature Matrices for Rare Cell Detection in Pediatric Cancer Liquid Biopsies

Kenia Viri[1]*, Uma K. Paithankar[2], Roma Parika[2], Sahbra Eldosougi[3], Uyen K. Ho[3], Caroline Zielinski[2], Sarah San Vicente[2], Jacy Fang[4], J.M. Shim[4], Timothy West[4], Jenna Barboza[1], Mason Perry[1], Sofia Neary[1], Isaac Rivera[1], Jackson Brenek[1], Jacqueline Tremblay[1], Lecia Sequist[5], Brian Nahed[4], Moshe Sade-Feldman[2], Shannon L. Stott[3], Jillian F. Wise[1]

[1]Department of Biology and Biomedical Sciences, Salve Regina University, Newport, RI 02840, [2]Krantz Family Center for Cancer Research, Massachusetts General Brigham, Boston, MA 02199, [3]Department of Biomedical Engineering, Boston University, Boston, MA 02215, [4]Department of Neurosurgery, Massachusetts General Brigham, Boston, MA 02199, [5]Early Detection and Diagnostics Program, Massachusetts General Brigham, Boston, MA 02199

Liquid biopsy is a minimally invasive approach for monitoring disease through rare circulating cells in blood. In pediatric brain cancer, detecting low frequency populations, including circulating tumor cells (CTCs), is challenging due to their scarcity and the complex cellular composition of blood. While single-cell RNA sequencing (scRNA-seq) provides high resolution characterization, its cost and time requirements make computational deconvolution an important approach for estimating cell type proportions. Our lab is developing a pipeline for pediatric liquid biopsy analysis by integrating optimized deconvolution with rare cell detection. To evaluate existing approaches, we generated pseudobulk datasets from scRNA-seq data from ten healthy pediatric blood samples. Each dataset contained varying proportions of twenty annotated cell types and sequential spike-ins of 0-100 vascular endothelial cells to simulate rare populations. We evaluated CIBERSORT, CIBERSORTx, Decon, ReDeconv, and DWLS using Pearson's correlation, root mean square error (RSME), and mean absolute deviation (mAD). CIBERSORTx Fractions demonstrated the strongest performance for predicting cells present at or below 0.03%, achieving a correlation of 0.52, while ReDeconv showed decreased correlation as rare cell proportions increased. All methods struggled to predict cell populations at 0%. To further optimize deconvolution, gene signature matrices were generated from healthy pediatric blood scRNA-seq datasets. Gene selection and normalization were evaluated using heatmaps and linear discriminant analysis (LDA), with log-normalized data achieving the highest accuracy compared with raw counts and variance stabilizing transformation (VST). These findings establish performance evaluations and optimize reference profiles for developing SparCell to improve rare CTC detection in pediatric biopsies.