Authors: Rachel Culp-Hill, Charles Nichols, Brendan Giles, Robert A. Law, Kian Behbakht, Benjamin G Bitler, Emma Crosbie, Vuna Fa, James Robert White, Abigail McElhinny

Category: CL01-03 Early detection biomarkers
Keywords: ovarian cancer, early diagnosis, multi-omics, biomarker

Title: Utilizing serum-derived lipidomics with protein biomarkers and machine learning for early detection of ovarian cancer in the symptomatic population

Ovarian cancer (OC) is the fifth leading cause of cancer-related deaths among women. Unfortunately, for most patients, detection of OC occurs at late stages (III/IV) when five-year survival is <30%. This is due in part to patients presenting with vague abdominal symptoms common in a variety of non-cancerous disorders (ex. gastrointestinal and gynecological conditions) that confound diagnosis. In addition, diagnostic methods for OC lack sensitivity and specificity for early-stage disease. Novel approaches and new biomarkers are urgently needed.

We conducted a multi-omics analysis of serum from two independent, clinically annotated biomarker discovery cohorts using UHPLC-MS untargeted lipidomics and a panel of protein biomarkers detected by manual immunoassay. Cohort #1 (N=544) specimens were obtained from the University of Colorado Gynecologic Tissue and Fluid Bank and commercial vendors. Samples included patients diagnosed with OC across subtypes and stages (N=219 total: 80 early-stage, 139 late-stage) and non-cancerous controls designed to mimic the symptomatic population. Controls included healthy donors (N=82), benign gynecological disorders (endometriosis, benign masses, etc. N=168), borderline tumors (N=25), and gastrointestinal disorders (irritable bowel syndrome, etc.: N=50). Cohort #2 (N=423) specimens were collected from a prospectively enrolled symptomatic population through Manchester University NHS Foundation Trust and supplemented with commercial vendor specimens. Samples included patients diagnosed with OC across subtypes and stages (N=109 total: 52 early-stage, 57 late-stage), individuals with benign gynecological disorders (N=86), borderline tumors (N=20), and symptomatic but otherwise normal individuals (N=208). Each cohort was processed independently.

Common lipid features were identified across the datasets to enable machine learning-based modeling on the individual and combined cohorts. Biomarker classes were modeled separately (lipids only, proteins only) and in combination (lipids and proteins), employing 20-fold cross validation. We consistently observed that a multi-omic model exhibits the highest AUC compared to individual biomarker classes. The 20-fold cross-validated AUC for the top-performing model applied to both cohorts was 95.0% (CI 93.5-96.5) for all controls vs. all OC, and 97.3% (CI 96.1-98.4) for all healthy vs. all OC.

In summary, LC-MS-based lipidomic profiling of serum combined with proteins represents a powerful diagnostic strategy, with multiple lipid species contributing to a diverse feature space. Combining multi-

omics and machine learning offers a promising new approach as a clinical diagnostic for detecting OC in this complex patient population. Future development efforts are aimed at narrowing biomarkers and validating performance in a prospectively collected cohort of symptomatic women.