A Novel Automated Analysis of Flow Cytometry Data Identifies Distinct Cell Signatures in the Peripheral Blood of CLL Patients — Preliminary Analysis

J. Rueter, MD1, V. Philip, PhD2*, K. Karuturi, PhD2*, Z. Oueida, MD1*, M. Chavaree, CTR1*, P. Helbig, CCRP1*, F. Eyerer1*, K. Mills, PhD2* and J. Graber, PhD2*

Introduction: The development of novel immunotherapeutic drugs for the treatment of hematologic malignancies has increased the need for novel biomarkers. To address this need, we developed a novel integrative analysis procedure for the automated analysis of multidimensional flow cytometry data obtained from the peripheral blood of patients with chronic lymphocytic leukemia (CLL). State of the art flow cytometry analysis is accomplished by manual sequential segmentation, or gating, of cell populations based on similarities in fluorescence and light scatter characteristics through visualization of the data in one- or two-dimensional plots. This approach has a number of limitations, including the subjective nature of the gating and the inability to fully utilize the high-dimensional data. Recent efforts have produced sophisticated computational methods that overcome many of these limitations; however, these newer computational methods have not been tested in a clinical context and have focused on the rigorous and automated analysis of samples from individual patients, with substantially less effort towards the analysis of patient populations. The ultimate goal of our analysis is to develop computational approaches that will enable an identification of subsets of patients with distinct immunological markers.

Methods: We used our analytic framework to analyze multidimensional flow cytometry data (26 cell surface markers in 4 different antibody cocktails) from peripheral blood specimens of a heterogeneous group of 55 CLL patients and 13 healthy controls. The markers that we chose for this analysis were based on known phenotypes of leukocyte subpopulations. We developed a novel analysis framework that facilitates automated identification of both common cell types and patient population subgroups, based on post-processing of individual sample analysis data with the FLOCK program. FLOCK identifies clusters of putatively similar cells in an individual sample by multidimensional clustering of the fluorescence marker and light-scattering measurements. We developed a rigorous hierarchical clustering approach to identify common “cell signatures” across multiple patients. The cell signatures were then mapped back onto the individual patient samples and used in a second clustering that identified patient subgroups based on similar abundances of specific cell types.

Results: We identified 99 cell signatures across our study population. These separated into 9 signature clusters. This allowed for 2 main patient clusters with further separation into 5 smaller sub-clusters. At least 2 of the signature clusters showed an enrichment of T cell-like signatures, with the rest of the clusters showing a mix of cell signatures, including T cells, B cells and neutrophils/other myeloid cells. The controls cluster tightly together and a few different patient clusters can be identified. The traditional clinical parameters of disease do not seem to correlate with these new patient clusters.

Conclusion/Discussion: Using a novel integrative analysis procedure to analyze complex flow cytometry data of the peripheral blood from CLL patients, we are able to identify distinct cell signatures which allow for the identification of specific subpopulations of patients and healthy controls.

Future directions: 1) Define optimal set of markers. 2) Refine parameters for cluster formation to establish an optimal number of signature clusters. 3) In order to establish a predictive biomarker, analyze a population that is enriched for patients that are about to start disease-modifying therapy.