Unlocking Biomarkers for Cancer Immunotherapy: Immune Monitoring with High-Parameter Flow Cytometry

Xiyu Peng¹, Fiona Ehrich¹, Ronglai Shen¹, Katherine S Panageas¹*

¹Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, NY, NY 10017, USA
*Correspondence should be addressed to Katherine S Panageas, panageak@MSKCC.ORG

Received date: May 01, 2024, Accepted date: June 06, 2024


Copyright: © 2024 Peng X, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Commentary

A decade ago, Science magazine named cancer immunotherapy as the breakthrough of the year, recognizing its pivotal clinical trials of antibodies designed to inhibit negative regulators of T cell function [1]. Antibody therapies targeting cytotoxic T lymphocyte antigen 4 (CTLA-4) and programmed cell protein 1 (PD-1) initially showed effectiveness against advanced melanoma and have since expanded in use to non-small cell lung, renal cell, and bladder cancers, among others [2-10]. In 2022, lymphocyte activation gene 3 protein (LAG-3) antibody, relatlimab, received FDA approval for unresectable or metastatic melanoma in combination with PD-1 blockade [11]. Despite the proven benefits of these agents, a large number of patients do not respond with positive outcomes. Among patients with previously untreated advanced melanoma, only approximately 30% and 40% achieve long-term disease control with first-line anti-PD-1 monotherapy and anti-PD-1 and anti-CTLA-4 combination therapy, respectively [12-14]. Among the subset of patients with anti-PD-1-refractory disease, only 30% respond to subsequent anti-PD-1 and anti-CTLA-4 combination therapy, respectively [12-14]. Among the subset of patients with anti-PD-1-refractory disease, only 30% respond to subsequent anti-PD-1 and anti-CTLA-4 combination therapy, respectively [15-16]. Furthermore, as the use of immunotherapy in clinical practice matures, the combination regimens that have emerged carry increased risk of toxicity [10]. Achieving the full potential of immunotherapy relies on tailoring treatments to individual patients to maximize effectiveness while minimizing side effects. The success of this endeavor depends on the identification of biomarkers.

Flow cytometry, an important tool in immunology, allows for detailed profiling of the various cell types that make up the immune system. This method generates complex and high-parameter data, characterizing immune cell populations that dynamically shift in response to treatments, and thus requires advanced analytical tools. The TopicFlow computational framework, developed by Peng and colleagues, meets these needs [17]. Utilizing Latent Dirichlet Allocation to analyze flow cytometry data, TopicFlow reveals the dynamics of T cell populations upon treatment exposure and identifies distinct immune response subgroups among patients, thereby deepening our understanding of the immune system’s response to immunotherapy. This detailed insight is essential for refining treatment approaches and improving patient outcomes.

TopicFlow is designed to uncover immunological topics within large-scale flow cytometry data from patients’ samples by conceptualizing cells as words, cell types as terms, and flow samples as documents (Figure 1). Each topic identified through TopicFlow is characterized by a probabilistic distribution across various cell types, delineating the ones that are likely to coexist and contribute to the same biological process. This method, in contrast to traditional approaches that assess cell types individually, allows for a simultaneous evaluation of all cell types and their interactions. By aggregating similar pharmacodynamic trends across multiple cell types, TopicFlow enhances the signal-to-noise ratio. Cell types that co-occur frequently and share similar pharmacodynamic changes tend to be more likely to be grouped into the same topic. In an unsupervised approach, TopicFlow identifies these biologically meaningful topics and provides estimates of the topic proportions for each patient sample. This allows for the stratification of patients into clusters based on the topic compositions. Further exploratory analysis can then be conducted to correlate patient subgroups to clinical outcomes enhancing the understanding of treatment impacts.
TopicFlow offers several advantages as a computational tool for analyzing flow cytometry data. Primarily, it is an innovative, statistically robust method that can be integrated seamlessly into existing single-cell analysis workflows. The input is a matrix of cell types by sample, derived from cell phenotyping procedures. For users interested in learning more details, we provide a comprehensive tutorial on our GitHub page (https://xiyupeng.github.io/LDA_examples/).

The study by Peng and colleagues illustrates how TopicFlow provides deep insights into highly complex flow cytometry datasets in immuno-oncology research [17]. In their study, they analyzed 138 peripheral blood samples, containing approximately 17 million T cells from 51 patients with melanoma treated with anti-PD-1 and anti-CTLA-4 combination therapy. These samples were collected at three time points: baseline pre-treatment (week 0) and on-treatment (week 3 and week 6). First, T cells were pooled across samples and classified into phenotypes via a graph-based clustering algorithm. TopicFlow was then applied to examine the dynamic changes in these cells, leading to the identification of three distinct T cell topics (Figure 2): (1) an activation topic, predominantly featuring memory T cells (Tcm/Tem), (2) a naive topic, mainly involving naive T cells; and (3) an exhaustion topic, comprised of terminally differentiated T cells (Temra). Each topic corresponds to different T cell phenotypes observed during treatment.

Further analysis of topic proportions across patients and over time recapitulated prior findings regarding treatment resistance and uncovered new insights into the dynamics of immune activation and immune-related toxicity. Earlier studies classified pre-treatment peripheral blood samples from patients into three “immunotypes,” including a “LAG+” immunotype characterized by a high abundance of LAG-3+ CD8+ T cells, which is associated with poorer treatment outcomes [18]. TopicFlow independently recognized a T cell exhaustion topic that aligns closely with the LAG+ immunotype and was present in high proportions in these patients’ samples. The clarity of the topic framework facilitated a deeper examination of the T cell functional characteristics within the LAG+ immunotype, providing opportunity to identify more precise and longitudinally stable biomarkers. The activation topic, which captures the co-expansion of multiple T cell populations including Ki67+ CD8+ T cells, emerged upon treatment, and serves as a complex pharmacodynamic marker of immune response. Relatedly, the naive T cell topic was found to decrease upon treatment, indicating T cell evolution as part of the overall immune response. Interestingly, a higher proportion of the naive topic at baseline was significantly associated with immune-related toxicity during treatment, suggesting its potential as an indicator of adverse immune reactions.

Currently, there are limited computational methods available capable of identifying pharmacodynamic biomarkers from molecular profiles in longitudinally collected patient blood samples. These methods face several challenges and gaps, including the need for integrating temporal data within highly complex single-cell datasets, managing missing data,
addressing unevenly spaced time points, and analyzing large-scale data across diverse patient cohorts. Most existing literature has focused on monitoring the pharmacodynamics of limited cell subsets, one at a time, such as the Ki67+ CD8+ T cell subset, longitudinally [19]. Previously Bachireddy et al. [20] utilized a Gaussian process regression model to study temporal dynamics of T cell states in a longitudinal single-cell RNA-seq dataset from a patient cohort undergoing cellular therapy. However, for flow cytometry, TopicFlow appears to be the first computational framework developed for this purpose. It has been used to analyze large-cohort, high-parameter flow cytometry data to identify biomarkers in peripheral blood samples. Moreover, the topic model framework can be extended to other types of single-cell datasets. For instance, we developed SpaTopic [21] for spatial single-cell datasets, integrating spatial information with Latent Dirichlet Allocation to identify tissue architecture from multiplexed imaging.

We intend to validate the promising scientific findings reported by Peng et al. with additional independent datasets. Furthermore, we plan to expand the use of TopicFlow to flow cytometry data from a variety of patient cohorts, encompassing a broader range of clinically relevant scenarios. TopicFlow is not limited to the analysis of flow cytometry data but can also be applied to analyze single-cell RNA-seq data, as demonstrated with a large-scale liver cancer single-cell RNA-seq dataset [22] in the tutorial. The original authors used a highly complex, domain knowledge-intensive method to assign each patient into one of the patient subtypes using ad hoc criteria. In contrast, TopicFlow is more streamlined and structured, requiring no prior knowledge input. It is a more de novo approach for the discovery of cellular subgroups among patients instead of being limited by existing knowledge.

A key finding, not discussed in the original study, was the identification of a patient subgroup characterized by a mixture of two cellular topics. Moreover, as an unsupervised technique, TopicFlow does not require a predefined expression profile for each cluster. Instead, it discerns patterns based solely on the pharmacodynamic changes and prevalence of cell subsets across samples. For instance, we highlight that T cell subsets under the same topic not only share similar pharmacodynamic trends, but also show analogous marker expressions [17].

**Figure 2:** Three topics (activation, naive, exhaustion) were identified by TopicFlow from a longitudinal flow cytometry dataset of melanoma patients treated with anti-PD-1 and anti-CTLA-4 combination therapy [17].
similarity suggests that these cell subsets are likely involved in similar biological processes during treatment, and grouping these cell types into topics helps clarify the pharmacodynamic profiles associated with specific treatments.

By combining TopicFlow with other computational tools currently under development in our group, we are confident in our capacity to identify blood-based biomarkers that predict outcomes or indicate the pharmacodynamics of various regimens in cancer immunotherapy. Ultimately, these biomarkers may be integrated into immune monitoring and hold potential to inform therapeutic strategies in clinical practice.

Acknowledgements

This work is supported in part by the MSKCC Society, the V Foundation, and the Parker Institute for Cancer Immunotherapy, NIH P30 CA008748, NIH R01 CA276286, and the MSK-MIND consortium.

References