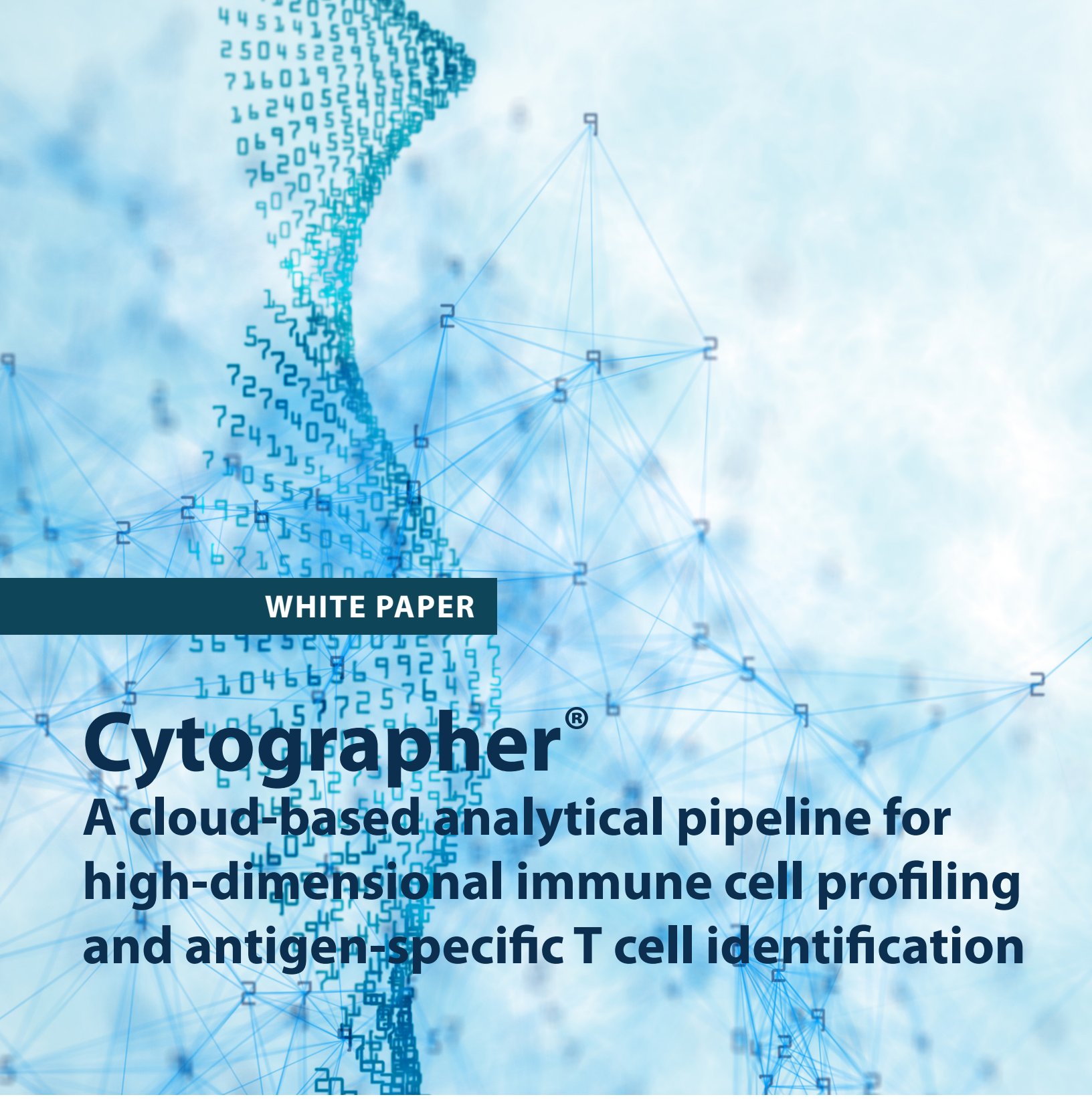

WHITE PAPER

Cytographer®

A cloud-based analytical pipeline for high-dimensional immune cell profiling and antigen-specific T cell identification

ImmunoScape

Cytographer®

A cloud-based analytical pipeline for high-dimensional immune cell profiling and antigen-specific T cell identification

Introduction

High-dimensional immune profiling and antigen-specific T cell identification has become indispensable for a comprehensive understanding of the immune response to cancer, infectious- and autoimmune diseases. It has the potential to generate comprehensive cellular data that may lead to the identification of biomarkers and inform the rationale for the development of therapeutic intervention strategies and vaccines.

Complex immune cell compositions in blood or tissue samples can be resolved using high-dimensional cell profiling through the identification and deep characterization of various immune cell subsets, including antigen-specific T cells, which are key effector cells in many innovative immunotherapy and vaccine strategies.

However, this approach generates a plethora of biological data that needs to be processed, analyzed and interpreted in the context of clinical parameters in order to derive meaningful insights. This white paper provides an overview of the Cytographer cloud-based analytical pipeline that allows custom high-dimensional immune cell profiling and antigen-specific T cell identification for providing in-depth insight into immunological data.

Overview

Designed and developed by ImmunoScape®, the proprietary cloud-based analysis platform Cytographer enables advanced analytics on multi-parametric mass- (CyTOF®) and flow cytometry data. Cytographer is securely accessible from anywhere in the world requiring only a web browser, and data are securely stored at a geolocation of the customer's choice in encrypted data vaults of Amazon AWS.

The platform offers a wide range of state-of-the-art analysis and antigen-specific T cell identification. The modular nature of the architecture using container technology (Docker) allows flexible scaling for processing of large datasets without compromising on speed and can seamlessly be upgraded with newly developed and published tools and algorithms (Figure 1).

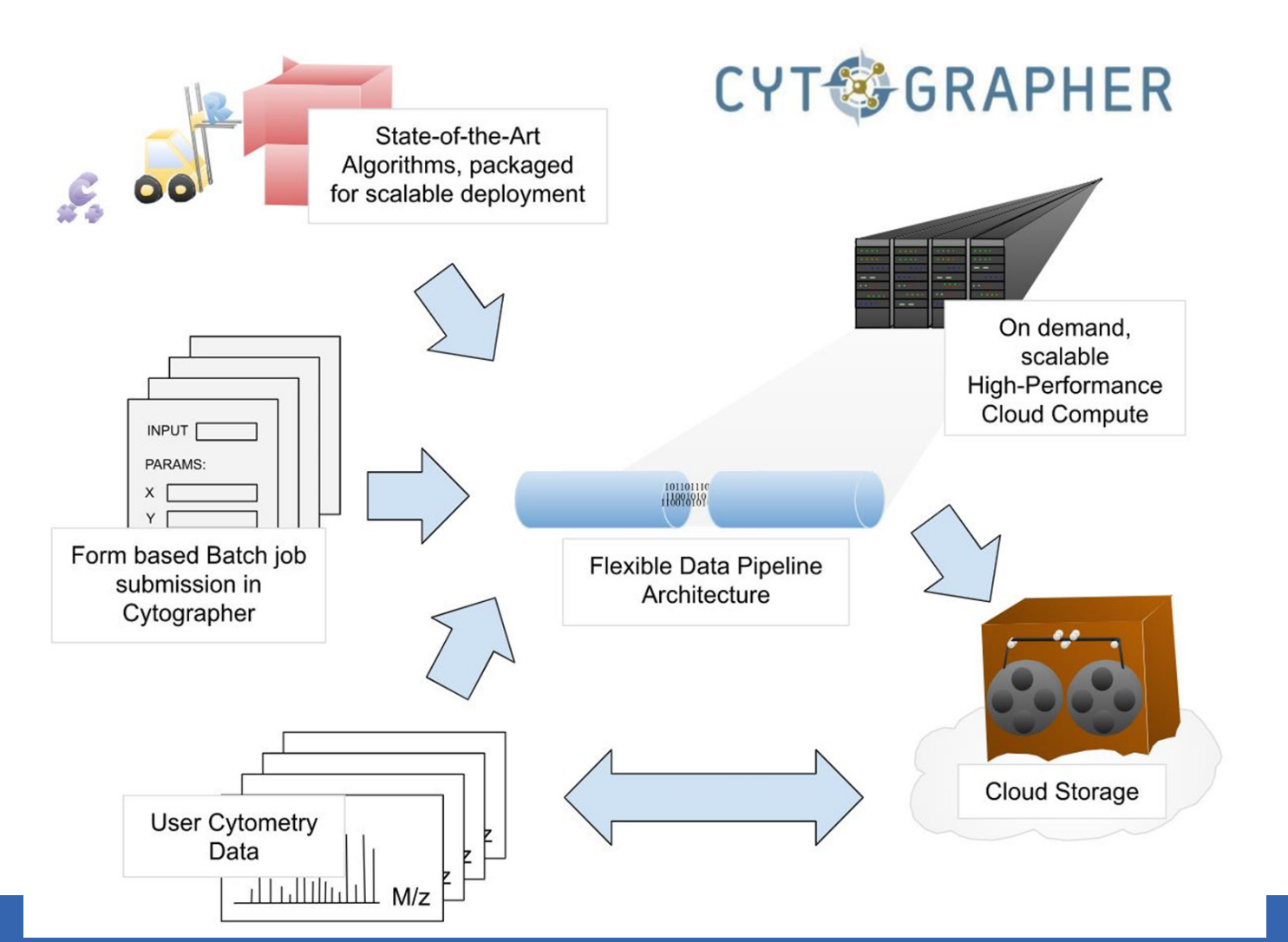

Figure 1. The underlying software architecture of the Cytographer platform supports scale-up without compromising on speed and can be seamlessly upgraded with newly developed and published tools and algorithms.

The intuitive and straightforward user interface of the Cytographer platform familiarizes users with conventional, as well as state-of-the-art high-dimensional data analysis tools, and facilitates complex data analysis without the need of programming experience or bioinformatic expertise. With a few clicks, a user can perform a series of standard data analyses or access more advanced tools to customize an analytical pipeline according to the study-specific experimental design and needs. The analytical pipelines generate graphical outputs for easy visual interpretation and/or data files for downstream processing. Data files can be downloaded or further analyzed in Cytographer.

Analytical Features

The Cytographer platform offers a wide range of analytical features for mass- and flow cytometry data which are categorized into different modules (Figure 2).

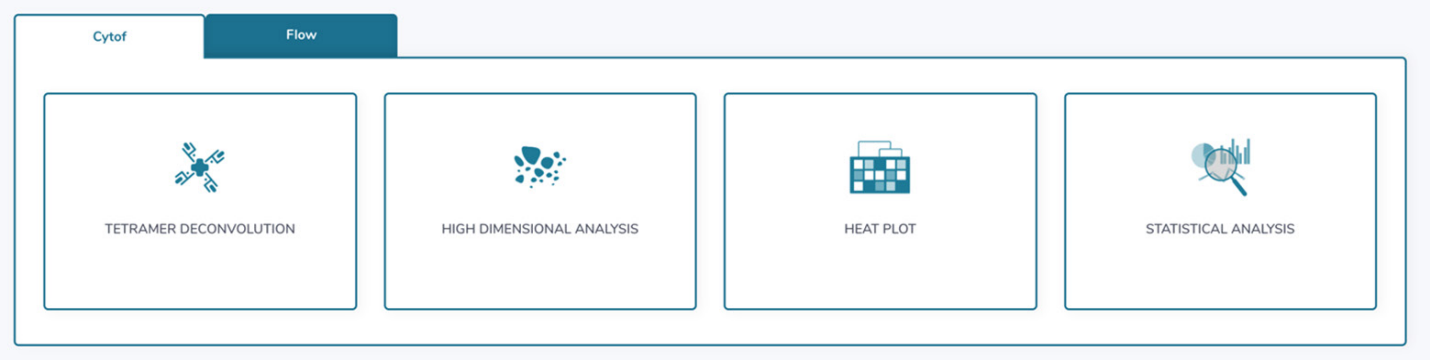

Figure 2. Analytical modules available in the Cytographer platform for mass and flow cytometry data

The high-dimensional analysis module includes several tools for high-dimensionality data reduction and cell clustering, coupled with graphical visualizations. The entire range of tools currently available in the Cytographer platform for assembly into custom analytical pipelines is listed in the Appendix. The Heat Plot and Statistical Analysis module allow the user to produce customized heat maps and to perform statistical analysis based on the input data file. Results can be downloaded as tables and are also highlighted in various graphical outputs such as box plots, bar plots, volcano plots and Principal Component Analysis (PCA) biplots.

ImmunoScape can identify rare antigen-specific CD8 T cells from a large screening of hundreds of epitope candidates in a single sample via the company's tetramer multiplexing methodology (TargetScape®) while still retaining the capacity to perform in-depth phenotypic profiling. The tetramer deconvolution module facilitates analysis of these multiplex data by providing tools for bona fide target identification including automated quantification and characterization of antigen-specific CD8 T cells.

Analytical Pipeline Execution

Each module allows the analysis of data stored and organized within a project workspace as well as data from previous analysis or instantly uploaded data files. Relevant analytical parameters, such as marker names are automatically extracted from the input data files.

The Cytographer platforms offers a standard- and advanced user mode where the latter unlocks the full range of modifiable parameters and configuration options. Methods can be executed individually, or consecutively as an analytical pipeline (Figure 3). When a pipeline is completed the user receives a notification email. The same analytical pipeline, whether default or custom designed, can be applied to multiple datasets to ensure consistent data processing across a study.

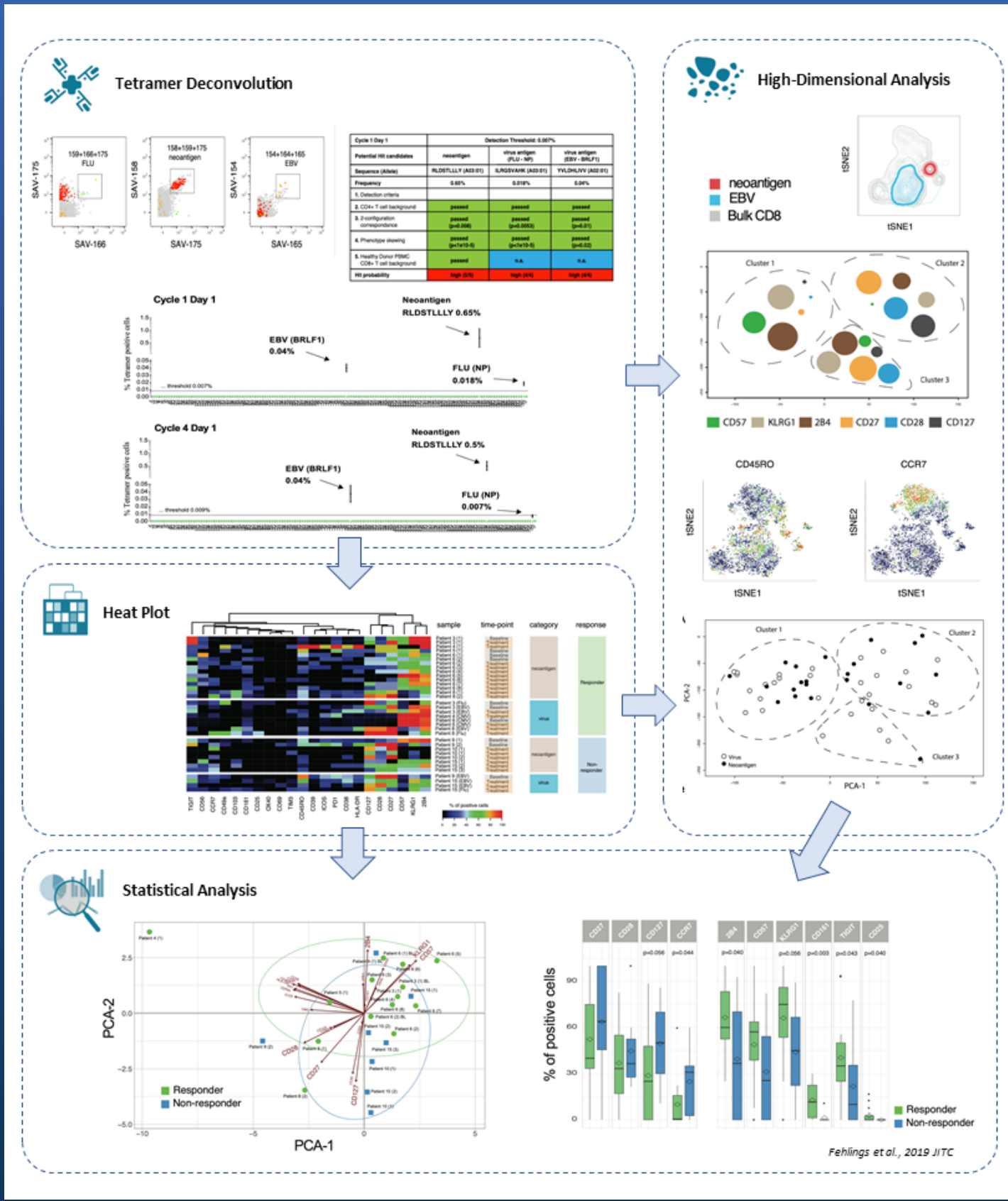

Figure 3. Example of an analytical pipeline executed in the Cytographer platform. PBMCs from a cohort of non-small cell lung cancer patients undergoing anti-PD-L1 treatment were screened for hundreds of tumor-specific mutant antigens (neoantigens) using tetramer multiplexing methodology (TargetScape). Bona fide neoantigen-specific T cells were identified using tetramer deconvolution module followed by high-dimensional phenotypic analysis of these cells. Subsequent statistical analysis deciphered differences between responder and non-responder patients. Results were visualized through several graphical outputs of each module. Data from Fehlings *et al.*, 2019 JITC.[1]

Available Options

Cytographer was initially developed to facilitate data analysis by the ImmunoScape team and is now available to the scientific community in the following manner:

The Analysis & Reporting Package is designed primarily for collaborators for whom ImmunoScape generates the data and analyzes the results. Full access is provided to the platform, as well as a dedicated in-house scientist responsible for data analysis and final reporting. The Cytographer platform's intuitive interface enables collaborators to perform additional data analyses on their own even after project conclusion.

The Access Package is intended for scientists who have independently generated mass or flow cytometry data. This user group can utilize Cytographer® tools for more advanced analytics for which ImmunoScape provides technical support.

Future Development

The analytical tools in the Cytographer platform include both open-source and proprietary software. Due to the modular architecture, users can expect continual addition of cutting-edge data analysis tools as soon as they become available. Ongoing developments include data pre-processing scripts as well as manual and automated gating tools in addition to sophisticated statistical methods for the detection of relevant immune cell subsets and their correlation with clinical outcomes.

Summary

The volume and complexity of data generated by high-dimensional immune profiling can provide valuable insight into how immunotherapy impacts the immune system. However, analysis of these datasets is complicated, tedious and typically requires higher-level bioinformatics skills to optimize the outputs.

The Cytographer proprietary cloud-based data analytic platform, empowers scientists by offering a one-source compilation of state-of-the-art bioinformatics tools for high-dimensional data analysis and antigen-specific T cell identification. The platform facilitates meaningful interpretation of complex biological datasets and their association with clinical parameters to enhance the potential value derived from clinical trials and other key studies.

Appendix: Analytical modules and tools

Tetramer Deconvolution
Automated peptide-MHC gating script for the detection of antigen-specific T cells (hits) using multiplexed tetramer technology for screening hundreds of epitope candidates in a single sample (TargetScape). Built-in statistical criteria facilitate an unbiased evaluation of bona fide antigen-specific T cells.

High-dimensional Analysis
Increased dimensionality of biological data requires tools for visualization and interpretable analysis. Dimensionality reduction methods that reduce the number of variables into few dimensions while preserving significant characteristics, and algorithms that group observations into discrete clusters are powerful tools for analyzing high-dimensional data. The Cytographer platform offers a wide array of techniques for data dimensionality reduction and cellular clustering which form the basis for population abundance analysis and biomarker discovery:

Principal component analysis (PCA)
PCA is a widely used and efficient technique for reducing dimensionality and visualizing relationships in multi-dimensional data.[2] PCA is able to segregate major populations of cells and to assess similarities and differences between samples by condensing data into a manageable number of summary variables (principal components). Classical PCA performs linear transformations which, however, precludes this technique for the segregation of rare or subtly different populations of cells in biological systems.

t-Distributed Stochastic Neighbor Embedding (t-SNE)
The algorithm in t-SNE reduces high-dimensional data down to two dimensions while preserving its local and global geometry.[3] t-SNE accounts for non-linear relationships between biological markers and maps closely related objects (such as similar cellular phenotypes) to nearby points in the two-dimensional space. It is widely used to visually delineate cell subsets and reveal the global structure of complex datasets. The method's strength is in detecting subtle variances in the overall phenotypes of cells.

Isometric feature mapping (ISOMAP)
ISOMAP measures geodesic distances between cells to learn the underlying global geometry between different cell types.[4] It computes a globally optimal solution to preserve nonlinear interactions and global relationships between cells and cell clusters. ISOMAP facilitates investigating differentiation trajectories and mapping phenotypic progressions between different clusters of cells.

Diffusion maps
Similar to ISOMAP, Diffusion maps preserve non-linear interactions and global relationships between cells.[5] The method allows for the detection of developmental trajectories and continuous branching events of differentiating cells including rare cell populations. Diffusion maps are mainly used for visualizing continuity in single-cell data with a robustness to noise and sample heterogeneity.

One-dimensional Soli-expression by Nonlinear Stochastic Embedding (One-SENSE)
Based on the t-SNE algorithm, One-SENSE allows the grouping of measured parameters into pre-defined categories.[6] In a One-SENSE plot, cells are projected onto a space composed of one dimension for each category, and each dimension (axis) represents one of these categories. Heat plots that are aligned in parallel to each axis allow for simultaneous visualization of each category across the plot and for direct assessment of the relationships between the categories. One-SENSE facilitates a type of categorical unsupervised analysis and can be used to resolve the relationships between the cellular arrangement and the underlying parameters

Uniform Manifold Approximation and Projection (UMAP)
UMAP is a novel manifold learning technique for dimensionality reduction that works similarly to t-SNE. As compared to t-SNE, UMAP has no computational restrictions in embedding large high-dimensional datasets and is often better at preserving global structures, thus revealing potentially more meaningful inter-cluster relations in the final projection. Shorter run times while providing good resolution of rare cell types and developmental trajectories make UMAP a valuable tool for a non-linear dimensionality reduction of single-cell data.

Potential of heat diffusion for affinity-based transition (PHATE)
PHATE is a dimensionality reduction method that preserves the local and global non-linear data structure using an information-geometric distance between data points.[7] PHATE provides a denoised visualization of high-dimensional data with many different underlying structures including trajectories, branches and clusters. Without imposing any strong assumptions on the structure of the data, PHATE is highly scalable in memory and run time and can be used to present large information into low dimensions.

PhenoGraph
PhenoGraph is an unsupervised clustering tool that partitions cells based on their connectivity to one another.[8] By creating a network that represents phenotypic similarities between cells, it facilitates the identification of sub-populations in high-dimensional single-cell data without prior knowledge. PhenoGraph runs efficiently on large datasets and enables coherent grouping of cells into meaningful populations with high stability.

FLowSOM
FlowSOM is an algorithm that uses self-organizing maps (SOM) as an unsupervised technique for cell clustering and visualization.[9] Utilizing hierarchical clustering, all events are sorted into a user-defined number of meta-clusters. Through successive iterations of training, each multidimensional data point is assigned to a node that it most closely resembles. FlowSOM shows high precision in preserving rare populations in distinct nodes.

Heat Plot and Statistical Analysis
Heat plots visualize various data inputs as color schemes in a two-dimensional representation. Differences in datapoints are represented by variations in the color intensities. Heat plots produced in Cytographer display marker Z-scores or non-scaled values and employ clustering using Euclidean distance or Pearson's correlation distances while still preserving sample grouping using distinguishable colors.

Cytographer® performs assumptions to test whether each data group is normally distributed and shows homogeneity of variance across levels. Depending on the assumptions made, different parametric or non-parametric tests are performed to test for significant differences between the input parameters. P-values are corrected for multiple testing using the False Discovery Rate (Benjamini-Hochberg) method. Data analyzed through Cytographer's® statistical module are visualized as PCA biplots, boxplots, bar plots and volcano plots, highlighting statistical differences in the input files.

References

1. Fehlings, M. et al. Late-differentiated effector neoantigen-specific CD8+ T cells are enriched in peripheral blood of non-small cell lung carcinoma patients responding to atezolizumab treatment. J. Immunother. Cancer 7, 249 (2019).
2. Ringnér, M. What is principal component analysis? Nat. Biotechnol. 26, 303–304 (2008).
3. Maaten, L. van der & Hinton, G. Visualizing Data using t-SNE. J. Mach. Learn. Res. 9, 2579–2605 (2008).
4. Tenenbaum, J. B., Silva, V. de & Langford, J. C. A Global Geometric Framework for Nonlinear Dimensionality Reduction. Science 290, 2319–2323 (2000).
5. Coifman, R. R. et al. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. Proc. Natl. Acad. Sci. 102, 7426–7431 (2005).
6. Cheng, Y., Wong, M. T., van der Maaten, L. & Newell, E. W. Categorical Analysis of Human T Cell Heterogeneity with One-Dimensional Soli-Expression by Nonlinear Stochastic Embedding. J. Immunol. Baltim. Md 1950 196, 924–932 (2016).
7. Moon, K. R. et al. Visualizing structure and transitions in high-dimensional biological data. Nat. Biotechnol. 37, 1482–1492 (2019).
8. Levine, J. H. et al. Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. Cell 162, 184–197 (2015).
9. Van Gassen, S. et al. FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. Cytom. Part J. Int. Soc. Anal. Cytol. 87, 636–645 (2015).

ImmunoScape

ImmunoScape, PTE. LTD.
1 Scotts Road #24-10
Singapore 228208

One Broadway, 14th Floor
Cambridge, MA 02142

immunoscape.com
info@immunoscape.com