Quantifying Cellular Heterogeneity in Cancer and the Microenvironment

By

Kirsten Elizabeth Diggins

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Cancer Biology

December, 2016

Nashville, Tennessee

Approved:

Jonathan M. Irish, Ph.D.

Vito Quaranta, M.D.

Melissa Skala, Ph.D.

Todd D. Giorgio, Ph.D.

ACKNOWLEDGEMENTS

This work would not have been possible without the financial support from the Vanderbilt-Ingram Cancer Center (VICC, P30 CA68485) and from the NIH/NCI (R25 CA136440-04). I am grateful to my advisor, Dr. Jonathan Irish, who has provided guidance and close mentorship through every step of my graduate career. I am also grateful to the mentorship and encouragement provided by the members of my dissertation committee: Dr. Vito Quaranta, Dr. Melissa Skala, and Dr. Todd Giorgio. Their guidance was instrumental in shaping the direction of my dissertation research and the success of my work.

In addition to my dissertation committee, I am also very grateful to Dr. Rebecca Ihrie and the current and past members of both the Irish and Ihrie labs. Their constant support as friends and colleagues as well as vital feedback about my research played a huge role in the success of work presented here.

I am deeply grateful to my family for their support. I thank my parents, Sarah and Rick Ansorge, for always encouraging me in my work and supporting my decisions, and my sister Rebecca Ansorge for being a constant friend and source of emotional support during my graduate school career. And finally, I am incredibly grateful to my husband, Zachary Diggins, for his constant support and encouragement. My success would not have been possible without him.

PREFACE

In this dissertation, I describe a set of methods for quantifying cellular heterogeneity and characterizing cell subpopulations in cancer and the microenvironment. Together, these methods establish a workflow for identifying biologically and clinically relevant cell subpopulations, quantifying the extent of cellular heterogeneity in biological samples, and quantitatively describing the features that are specifically enriched on cell populations. The work described here that has been submitted to or published in peer reviewed scientific journals is inserted into the body of the dissertation in the same form in which it was submitted for publication (Chapters 2-4). Many talented colleagues and collaborators played a major role in the development and success of this work. My collaboration with Dr. Melissa Skala's lab at Vanderbilt University, in which I worked closely with former graduate student Amy Shah, merits special mention here. The results of this collaboration are described in Appendix A.

Chapter 1 introduces the central goals and hypotheses of this dissertation. The biological and clinical significance of cellular heterogeneity in cancer is briefly introduced, followed by a short introduction to flow and mass cytometry and single-cell data analysis tools. These tools are further discussed in the context of the modular workflow explained in Chapter 2, as the foundation for the cellular heterogeneity measure described in Chapter 3, and as a method for analyzing single-cell imaging data (Appendix A). Novel methodological approaches for standardizing and automating single-cell data analysis are introduced. These methods are developed throughout the following chapters of the thesis and further discussed in Chapter 5. Specific topics relevant to each chapter are included in that chapter's preface.

Chapter 2, entitled "Methods for discovery and characterization of cell subsets in high dimensional mass cytometry data," describes a novel modular workflow that incorporates multiple computational tools, including biaxial gating, t-distributed stochastic neighbor embedding (tSNE), and spanning-tree progression analysis of density-normalized events (SPADE) analysis into a workflow that facilitates discovery of both abundant and rare cell populations in single-cell data. The phenotypic signatures of the identified cell populations are visualized in a clustered heatmap. This workflow is presented and demonstrated through a use case in which AML blast, non-blast, and normal human bone marrow cell subpopulations were identified and characterized after multi-dimensional mass cytometry analysis. This work provides a scalable and automated analysis pipeline that can

be applied to multi-dimensional, single-cell datasets arising from experimental platforms including mass cytometry and multiplexed flow cytometry.

Chapter 3, entitled “High-Dimensional Analysis of Acute Myeloid Leukemia Reveals Phenotypic Changes in Persistent Cells during Induction Therapy,” describes a method for quantifying longitudinal, phenotypic heterogeneity of acute myeloid leukemia cell subpopulations over the course of treatment. The 27 antibodies measured on AML patient samples by mass cytometry were used for dimensionality reduction analysis (tSNE), and the cells were displayed in the resulting two-dimensional phenotypic space. The distances between AML and normal cell populations in this reduced-dimensional coordinate space were used as quantitative representations of immunophenotype difference. Persistent (i.e. therapy resistant) cells were found to be more phenotypically distinct from normal hematopoietic stem cells (HSCs) compared to non-persistent AML or normal cell subpopulations. This work provides a conceptual framework that can be expanded and developed into additional tools for quantifying cellular heterogeneity from single-cell data in disease and normal biological contexts.

Chapter 4, entitled “Characterizing cell subsets in heterogeneous tissues using marker enrichment modeling,” describes a novel method termed Marker Enrichment Modeling (MEM) that automatically creates cell population labels that quantify population-specific feature enrichment. The concept and equation behind MEM analysis is introduced and demonstrated through several use cases, including analysis of normal human blood, bone marrow, and tonsil samples; murine myeloid cells; human glioblastoma samples; and human melanoma samples. A measure of population similarity is then proposed that calculates the pairwise difference between cell population MEM labels, enabling direct comparison of cell populations from different experiments, institutions, and platforms. MEM is therefore presented as a first step towards fully automated machine learning cell identity, a concept elaborated upon in Chapter 5.

Chapter 5 discusses the findings and approaches described in Chapters 2-4 and elaborates on future directions of this work, specifically in terms of machine learning cell identity using MEM labels and adapting these analysis tools for clinical applications.

Appendix A describes the application of tSNE analysis for visualizing the results of *in vivo* autofluorescence imaging. NAD(P)H and FAD, two cellular metabolites that are commonly dysregulated in cancer, were imaged

by optical metabolic imaging in squamous cell carcinoma tumors were implanted into mice and subsequently treated with common chemotherapeutic compounds. The optical metabolic imaging parameters were used in a tSNE analysis to visualize the individual cells in multi-parametric space. The analysis revealed that tumors treated with cetuximab or cisplatin displayed a greater degree of metabolic heterogeneity than the control group. This work therefore demonstrated that single-cell tools such as tSNE can be applied for data analysis in various experimental platforms.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

CHAPTER 1

INTRODUCTION

Biological and clinical significance of cellular heterogeneity in cancer

Tumors are composed of heterogeneous populations of cancer cells that undergo rounds of clonal selection and expansion in the presence of microenvironmental influences (Hanahan and Weinberg 2011). Mutations and genomic instability, combined with evolutionary pressure from surrounding stromal and immune cells, gives rise to multiple cancer cell populations with distinct phenotypes and functions. Genetic changes to oncogenes or tumor suppressor genes, considered “driver” mutations, enhance the cell’s proliferative potential and ability to avoid both internal and external death signals. Cancer cells can resist or adapt to chemotherapy by acquiring novel driver mutations or losing expression of drug-targeted proteins. Therapies targeted against specific mutations in cancers therefore frequently fail to cure the patient because the cells ultimately evade the therapy through clonal selection.

Intratumoral heterogeneity is thus highly relevant to prognostic and therapeutic decisions in cancer. Clonal heterogeneity has been found to correlate with prognosis and therapy response in many types of cancer. For example, high clonal heterogeneity was found to correlate with increased risk of progression in esophageal cancer patients (Maley, Galipeau et al. 2006). The presence of subclonal driver mutations was also predictive of rapid progression in chronic lymphocytic leukemia (Landau, Carter et al. 2013). This correlation between high degrees of cellular heterogeneity and poor outcome has been noted across cancer types, including but not limited to head and neck squamous cell carcinoma (Mroz, Tward et al. 2013), endometrial cancer (Supernat, Lapinska-Szumczyk et al. 2014), and acute myeloid leukemia (Baer, Stewart et al. 2001, Patel, Gonen et al. 2012).

The immune system, known to be composed of highly heterogeneous and plastic cell populations, also plays a major role in cancer promotion, therapy response, and patient outcomes. For example, a heterogeneous subset of myeloid cells termed myeloid derived suppressor cells (MDSCs) is expanded in many types of cancer. This immature cell subset secretes factors to promote a malignant microenvironment and alter the immune

response to cancer (Berraondo, Minute et al. 2016). Tumor infiltrating lymphocytes (TILs) have also been shown to be a highly heterogeneous subset of immune cells (Hadrup, Donia et al. 2013). As a mechanism of immune evasion, cancer cells can suppress the anti-cancer activity of TILs, and as a result immunotherapy fail (Smyth, Ngiow et al. 2016). Measuring heterogeneity of tumor-associated immune cells is therefore crucial to understanding the immune response to cancer and continuing to develop and improve immunotherapy for cancer.

Single-cell versus aggregate analyses of human tissue

Traditional analyses of cancer average the genetic, phenotypic, and functional features of all the cells in a tumor. Previous efforts to quantify tumor heterogeneity have focused on deep sequencing analysis and the inference of cell subsets based on varying allele frequencies (Zare, Wang et al. 2014). However, protein levels do not always correlate with transcript levels due to post-translational modification and regulatory degradation, making it beneficial to measure protein expression level and activation in order to gain a complete understanding of cellular phenotype and function (Vogel and Marcotte 2012). Going forward, a quantitative, single-cell approach to characterizing intratumoral heterogeneity will aid in the development of novel therapeutic strategies that address the high rates of therapy resistance and relapse that is likely caused by the presence of multiple cell subpopulations in a given patient's cancer.

In recent years, several new platforms have been developed for measuring features at the single-cell level. Mass cytometry (CyTOF) allows for more than 30 features to be measured simultaneously at the single-cell level (Dalerba, Dylla et al. 2007, Prince, Sivanandan et al. 2007) (Figure 1-1A). Phospho-flow cytometry and multi-parametric immunophenotyping approaches have been successfully employed in the field of cancer, for example, to identify minimal residual disease in AML (Amir el, Davis et al. 2013), to find signaling node activation states in AML that predict response (Irish, Hovland et al. 2004), and to identify a negatively prognostic cell subset characterized by its altered signaling in follicular lymphoma (Irish, Myklebust et al. 2010) (Figure 1-1B). Other experimental platforms are also being developed for single-cell assays. Single-cell RNA sequencing is now being used for applications such as measuring cell-level genetic changes in glioblastoma cells (Patel, Tirosh et al. 2014). Multiplexed imaging mass cytometry can now be used to assess protein changes at the single-cell level

in tissues while retaining information about tissue structure and cell localization (Giesen, Wang et al. 2014). As these tools become increasingly high-throughput, novel computational data analysis tools will be needed to parse the large datasets.

Single-cell data analysis

Flow cytometry experiments yield data in the form of flow cytometry standard (FCS) files, in which per-cell feature expression is recorded in an *n* cells by *m* parameters matrix. The standard method for analyzing this data is biaxial gating, a method in which the analyst manually groups populations of cells based on their co-expression of pairs of measured markers. However, the widespread use of multi-dimensional flow cytometry has driven the development of computational tools and algorithms to facilitate cell subset discovery in high-dimensional datasets where pairwise biaxial analysis would be impractical or insufficient for identifying phenotypically distinct cell populations. Clustering tools like spanning-tree progression analysis of density-normalized events (SPADE) are frequently used for analysis of high-dimensional flow cytometry data, for instance, to cluster and visualize immune cells along a hematopoietic development trajectory (Bendall, Simonds et al. 2011, Qiu, Simonds et al. 2011, Bendall, Davis et al. 2014). Dimensionality reduction tools have also been adapted for use with flow cytometry data, such as t-distributed stochastic neighbor embedding (tSNE). The tSNE algorithm has been used to identify subsets of immune cells and leukemic cells (Amir el, Davis et al. 2013), identify and characterize understudied populations of murine myeloid cells (Becher, Schlitzer et al. 2014), and characterize the diversity of tissue-specific T cells (Wong, Ong et al. 2016), among other applications. Both SPADE and tSNE are discussed in further detail in Chapter 2. Many other computational tools have been developed for use with single cells data, including variance maximization (Newell, Sigal et al. 2012), mixture modeling (Pyne, Hu et al. 2009, Mosmann, Naim et al. 2014, Naim, Datta et al. 2014), spectral clustering (Zare, Shooshtari et al. 2010), and density-based automated gating (Qian, Wei et al. 2010) (Figure 1-1C).

Once identified, populations of cells are typically compared univariately by quantifying differences in median expression levels of measured proteins or other cellular features. Statistical methods are sometimes applied for comparing flow cytometry distributions, including variations on area under the curve (AUC) analysis (Kim, Donnenberg et al. 2016), K-S statistic-based methods (Cox, Reeder et al. 1988), and z-score calculations.

However, many of these types of tools make assumptions about distribution shape; for example, z-score assumes a normal distribution, and flow cytometry parameters are rarely normally distributed. Distribution differences also do not necessarily reflect population-specific enrichment of the measured feature. These methods also fail to capture both intra-population and inter-population variance in a single value.

While useful for population identification and univariate comparisons of distributions, the available analysis tools do not provide a quantification of marker enrichment, quantify cellular heterogeneity based on phenotypic differences, or provide adequate means by which populations can be registered and identified between samples in an unsupervised manner.

Addressing the need for novel analysis methods

This work describes novel workflows and tools to quantify cellular heterogeneity and determine population-specific feature enrichment, while laying the groundwork for further work in machine learning applications to automate cell population discovery and characterization. Chapter 2 establishes a workflow for identifying both rare and abundant cell populations in multi-dimensional flow cytometry datasets. This modular workflow uses biaxial gating for quality control gating, viSNE for dimensionality reduction and major cell population gating, followed by SPADE analysis for automated clustering. The results are visualized as a hierarchically clustered heatmap to give a comprehensive, qualitative view of population-level phenotypic signatures. In Chapter 3, a method is described for quantifying cellular heterogeneity based on multi-dimensional phenotypic distances within the tSNE reduced-dimension coordinate space. Using this method, the phenotypic differences that can be visualized and qualitatively interpreted using the workflow described in Chapter 1 can additionally be quantified, statistically analyzed, and correlated with clinical parameters. Finally, Chapter 4 describes a method for quantifying population-specific feature enrichment in the context of the sample, a set of samples, or compared to other specific cell subsets. These tools and workflows for quantifying cellular heterogeneity are applied in the analysis of normal human immune cells, AML, and other types of cancer in Chapters 2-4. Together, the work presented here provides novel methods for quantifying cellular heterogeneity in normal human tissue development, cancer, and other diseases.

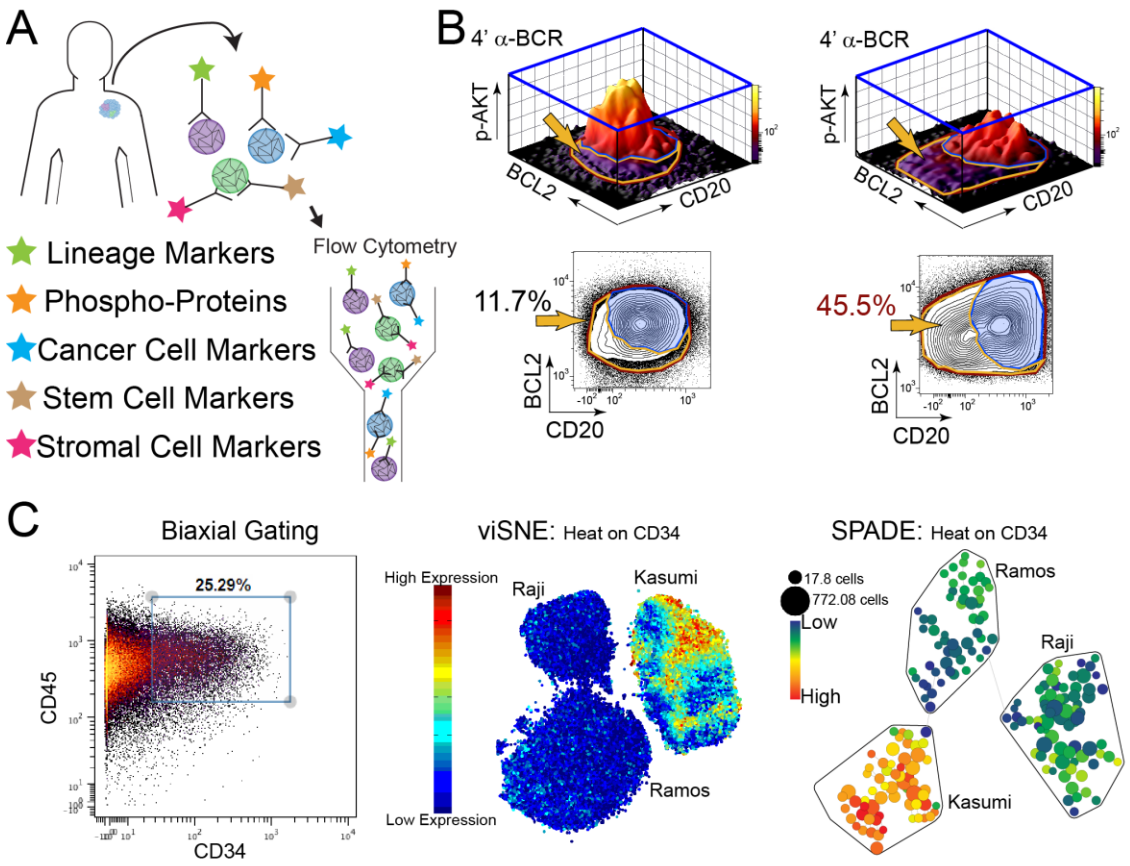

Figure 1-1. High dimensional phospho-flow and mass cytometry allows for identification and characterization of distinct cell subpopulations. A) Primary human tissue samples can be analyzed by high dimensional flow cytometry to measure clinically relevant markers and phosphorylation events. Stars represent different categories of markers that can be measured. B) Negative prognostic cell subsets in follicular lymphoma were deficient in signaling through BCR. The patient with a better outcome had a smaller proportion of LNP cells (left) compared to the patient with a worse outcome (right) (Irish, Myklebust et al. 2010). C) Biaxial gating, viSNE, and SPADE are three tools that can be used to identify populations of cells. Surface markers on three cell lines were measured by CyTOF. SPADE and viSNE were used for dimensionality reduction and subsequent gating. Raij and Ramos = Burkitt's lymphoma cell lines, Kasumi = AML cell line.

CHAPTER 2

METHODS FOR DISCOVERY AND CHARACTERIZATION OF CELL SUBSETS IN HIGH DIMENSIONAL MASS CYTOMETRY DATA

Authors: Kirsten E. Diggins, P. Brent Ferrell, Jr., and Jonathan M. Irish

This work is presented as it appears in manuscript form in *Methods* 2015 (Diggins, Ferrell et al. 2015).
http://www.sciencedirect.com/science/article/pii/S1046202315001991

Preface

The high dimensional data resulting from single-cell experimental technologies cannot be sufficiently interpreted using traditional gating strategies alone. I therefore sought to develop a computational data analysis workflow that would retain the benefits of biaxial gating and expert guided analysis while also automatically grouping phenotypically distinct cell subsets in high dimensional space. To achieve this goal, I combined manual gating, dimensionality reduction with viSNE, cluster analysis using SPADE, and clustered heatmaps in order to identify and visualize the phenotypic signatures of cell subpopulations from normal human bone marrow and AML patient blood. Workflows like the one presented here are essential to expedite and standardize clinical and basic biology studies that aims to identify and track relevant cell subpopulations between conditions, over time, and between patients. The work presented here describes the first step in my overall goal of standardizing the methods by which populations of cells are identified and then quantitatively labeled.

Abstract

The flood of high-dimensional data resulting from mass cytometry experiments that measure more than 40 features of individual cells has stimulated creation of new single cell computational biology tools. These tools draw on advances in the field of machine learning to capture multi-parametric relationships and reveal cells that are easily overlooked in traditional analysis. Here, we introduce a workflow for high dimensional mass cytometry data that emphasizes unsupervised approaches and visualizes data in both single cell and population level views. This workflow includes three central components that are common across mass cytometry analysis

approaches: 1) distinguishing initial populations, 2) revealing cell subsets, and 3) characterizing subset features. In the implementation described here, viSNE, SPADE, and heatmaps were used sequentially to comprehensively characterize and compare healthy and malignant human tissue samples. The use of multiple methods helps provide a comprehensive view of results, and the largely unsupervised workflow facilitates automation and helps researchers avoid missing cell populations with unusual or unexpected phenotypes. Together, these methods develop a framework for future machine learning of cell identity.

Introduction

High dimensional single cell biology

Single cell biology is transforming our understanding of the biological mechanisms driving human diseases and healthy tissue development (Irish and Doxie 2014). Mass cytometry is a recently developed technology that enables simultaneous detection of more than 40 features on individual cells (Bandura, Baranov et al. 2009, Ornatsky, Bandura et al. 2010). High dimensional mass cytometry measurements are single cell, quantitative, and well-suited to unsupervised computational analysis. New analysis tools have been created to take advantage of the massive amounts of data that result from high content single cell techniques like mass cytometry. Variations of many of these tools have been developed and applied for gene expression analysis, a field facing similar problems with data dimensionality. These tools draw on advances in machine learning and statistics that are not yet widely applied in biological studies. Many of these tools are complementary and address different aspects of data analysis, and it can be challenging for biologists to know when and how to use these tools to get the most out of their data. Advances have also been made in automating and standardizing the flow cytometry data analysis workflow (Pyne, Hu et al. 2009, Aghaeepour, Finak et al. 2013, Finak, Frelinger et al. 2014). Here, we present a modular workflow focused on high dimensional single cell analysis that combines multiple tools to provide a comprehensive view of both cells and populations. Rather than making the workflow fully automated, the goal here was to combine the complementary benefits of expert analysis and machine learning. This approach maintains single cell views, provides automatic population assignment for each cell, and facilitates statistical comparison of the key cellular features that characterized each population. This semi-supervised workflow facilitates comparison of populations discovered by different computational

approaches, in different clinical samples, or using different biological features (e.g. RNA expression, cell surface protein expression, and cell signaling).

An advantage of traditional analysis in flow cytometry is the reliance on identification of known, prominent populations with strong supporting biology in the literature. Given the typical panel size for fluorescent experiments, this type of supervised analysis is fast and usually adequate. Unfortunately, expert manual gating has been shown to be particularly prone to inter-operator variability (Maecker, Rinfret et al. 2005) and a tendency to overlook cell populations (Krutzik, Clutter et al. 2005, Amir el, Davis et al. 2013, Irish 2014). Recent efforts have developed new tools for high dimensional cytometry data that bring in elements of machine learning and statistical analysis, including clustering (Sugar and Sealfon 2010, Bendall, Simonds et al. 2011, Qiu, Simonds et al. 2011, Bruggner, Bodenmiller et al. 2014), dimensionality reduction (Amir el, Davis et al. 2013), variance maximization (Newell, Sigal et al. 2012), mixture modeling (Pyne, Hu et al. 2009, Mosmann, Naim et al. 2014, Naim, Datta et al. 2014, Chen, Hasan et al. 2015), spectral clustering (Zare, Shooshtari et al. 2010), neural networks (Tong, Ball et al. 2015), and density-based automated gating (Qian, Wei et al. 2010). Here, we highlight use of these tools in a sequential single cell bioinformatics workflow (Table 2-1). In particular, different tools address aspects of data visualization, dimensionality reduction, population discovery, and feature comparison. It can be valuable to apply multiple tools in order to view data in different ways and fully extract biological meaning at the single cell level (Figure 2-1) and the population level (Figure 2- 2 and Figure 2-3). After identifying cell subsets with the aid of computational tools, measured features, such as protein expression in the examples here, can be compared between and within the subsets. Traditional statistics used include medians, variance, and fold changes. Other statistical methods such as histogram statistics and probability binning have also been used to compare distributions in flow cytometry data (Bagwell, Hudson et al. 1979, Overton 1988, Roederer, Moore et al. 2001).

Table 2-1 – A modular machine learning workflow for semi-supervised high-dimensional single cell data analysis

Analysis step		Traditional	Additional methods[§]	Method here
Data collection	1) Panel design	Human expert	-	-
	2) Data collection	Human expert	-	-
Data processing	3) Cell event parsing	Instrument software	Bead normalization and event parsing (Finck, Simonds et al. 2013)	-
	4) Scale transformation	Human expert	Logicle (Moore and Parks 2012)	-
Distinguishing initial populations	5) Live single cell gating 6) Focal population gating	Biaxial gating + human expert	No event restriction, AutoGate (Meehan, Walther et al. 2014)	viSNE + human expert (Figure 1)[†]
Revealing cell subsets	7) Select features	Human expert	Statistical threshold (Irish, Hovland et al. 2004)	Human expert[†]
	8) Reduce dimensions or transform data	N/A	Heat plots (Irish, Myklebust et al. 2010), SPADE (Qiu, Simonds et al. 2011), t-SNE (Geoffrey Hinton 2002), viSNE (Amir el, Davis et al. 2013), ISOMAP (Tenenbaum, de Silva et al. 2000), LLE (Roweis and Saul 2000), PCA in R/flowCore (Hahne, LeMeur et al. 2009)	SPADE[†], viSNE
	9) Identify clusters of cells	Human expert	SPADE, k-medians, R/flowCore, flowSOM (Van Gassen, Callebaut et al. 2015), Misty Mountain (Sugar and Sealfon 2010), JCM (Pyne, Lee et al. 2014), ACCSENSE (Shekhar, Brodin et al. 2014), DensVM (Becher, Schlitzer et al. 2014), AutoGate, *Citrus* (Bruggner, Bodenmiller et al. 2014)	SPADE (Figure 2)[†], viSNE + human expert (Figure 1)
	10) Cluster refinement	Human expert	Citrus, DensVM, R/flowCore	-
Characterizing cell subsets	11) Feature comparison	Select biaxial single cell views	viSNE, SPADE, Heatmaps (Irish, Hovland et al. 2004, Kotecha, Krutzik et al. 2010), Histogram overlays (Irish, Hovland et al. 2004, Kotecha, Krutzik et al. 2010), Violin or box and whiskers plots (Hahne, LeMeur et al. 2009), Wanderlust (Bendall, Davis et al. 2014), Gemstone	Heatmaps (Figure 3A)[†], viSNE (Figure 1C), SPADE (Figure 2C)
	12) Model populations	N/A	Median (Irish, Hovland et al. 2004), JCM, PCA	-
	13) Learn cell identity	Human expert	-	Human expert[†] (Figure 1B, Figure 2B, and Figure 3B)
	14) Statistical testing	Prism, Excel	R/flowCore	-

[§]Methods with broad application (e.g. R/flowCore) are listed minimally at select steps based on particular strengths or published applications.
[†]Denotes the primary approach used at each step in the sequential analysis workflow shown here.

Overview of the analysis workflow

The workflow presented here was applied to a CyTOF dataset from the analysis of healthy human bone marrow and a diagnostic sample of blood from a patient with acute myeloid leukemia (AML). The annotated FCS files and a step-by-step guide are available online from Cytobank (www.cytobank.org/irishlab) (Kotecha, Krutzik et al. 2010) and FlowRepository (http://flowrepository.org/experiments/640) (Spidlen, Breuer et al. 2012). This workflow was developed for use with high-dimensional mass cytometry data. However, it can also be applied to fluorescent flow cytometry data. The main steps presented consist of event restriction, population discovery, and population characterization. Each of these aspects of data analysis can be achieved with a variety of techniques (Table 2-1), and some tools address multiple steps. By sequentially combining three different techniques, this workflow draws on the strengths of specific tools, keeps biologists in touch with single cell views, and enables analysis of data from different studies and single cell platforms.

In the case of the example dataset here, the overall biological goal was to identify and compare three populations of cells: leukemia cells (AML blasts) and non-malignant cells (non-blasts) in the blood of a leukemia patient, and bone marrow cells from a healthy donor. In the analysis workflow, cell events were first manually gated based on event length and DNA content to include intact, single cells (Figure 2-1) (Bendall, Simonds et al. 2011). Next, visualization of stochastic neighbor embedding (viSNE) was used to identify and gate major subsets (Figure 2-1). Gated cells from healthy bone marrow and AML were then analyzed by spanning-tree progression analysis of density-normalized events (SPADE) to discover and compare cell subsets (Figure 2-2). Finally, the cell subsets identified by SPADE were further characterized using complete linkage hierarchical clustering and a heatmap in R (Figure 2-3). The details of mass cytometry data collection and processing prior to initial cell selection (gating) are not covered in detail here. These early steps include experiment design, collection of data at the instrument (and instrument setup), any normalization, and transformation of the data to an appropriate scale (Table 2-1).

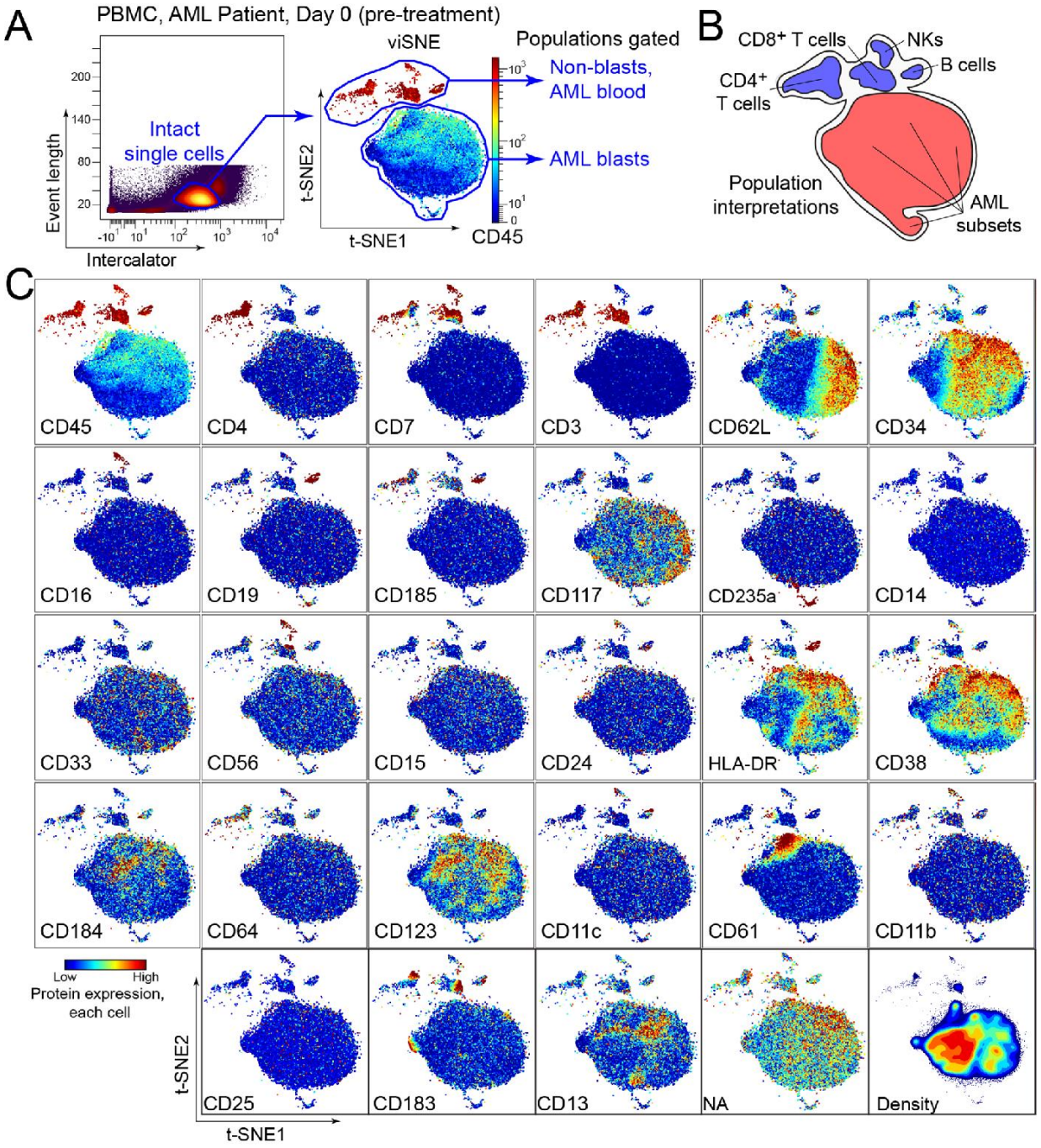

Fig. 2-1. Distinguishing initial populations with viSNE analysis of per-cell protein expression and expert gating. Plots show the use of viSNE to obtain a comprehensive single cell view and to initially distinguish cancerous and non-malignant cells in the blood of an AML patient. A) Expert analysis of mass cytometry data identified intact single cells using event length and intercalator uptake. Subsequent viSNE analysis arranged cells along unitless t-SNE axes according to per-cell expression of 27 proteins. Expression of CD45 protein is shown for each cell on a heat scale. viSNE automatically arranged leukemia cells in one area of the map and facilitated selection of AML blast and non-blast cells by expert gating. Populations identified by viSNE and expert gating were subsequently analyzed by SPADE (Fig. 2-2). B) Human interpretation of population identities based on viSNE analysis is shown. C) Plots show expression of the 27 proteins, nucleic acid intercalator (NA), and density measured per cell.

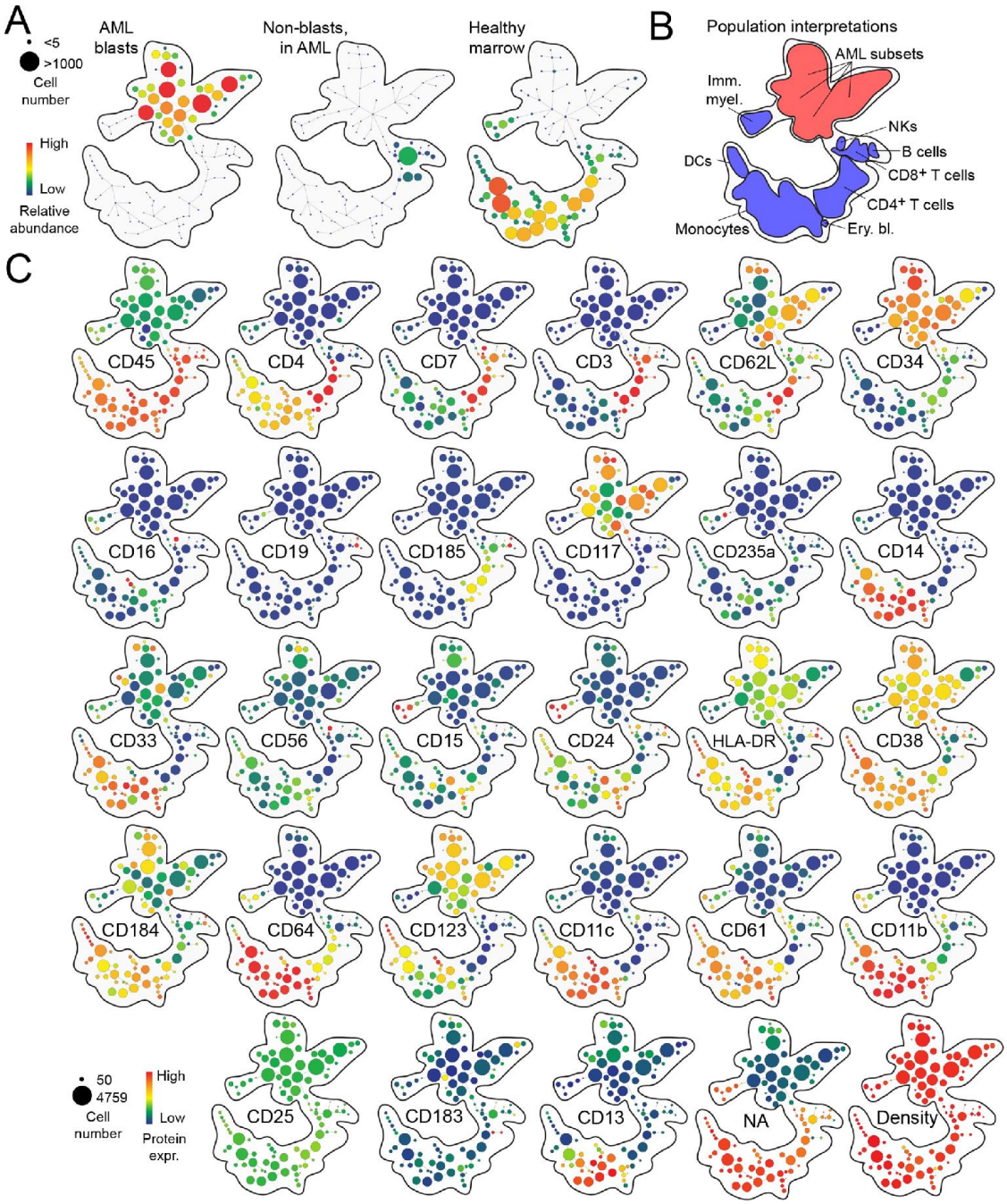

Figure 2-2. Revealing cell subsets with SPADE analysis of population hierarchy, cell abundance, and median protein expression. Plots show the use of SPADE to reveal clusters of cell subsets in cell populations identified by expert analysis and viSNE (Fig. 1). A) SPADE analysis identified distinct population clusters in each sample. Cell abundance is represented by size and color of each circle representing a population of cells. Phenotypically distinct cell subsets fell into different regions of the SPADE tree. B) Human interpretation of population identities based on SPADE analysis is shown. C) Plots show expression of the 27 proteins, nucleic acid interalator (NA), and density measured per cell**.**

The initial event restriction step that begins the workflow focuses the analysis on populations of cells. The goal at this step is to remove events that do not contribute useful information while making minimal changes to the data and not over-focusing. Event restriction is traditionally performed using biaxial gating (Table 2-1), but given

the high dimensionality of mass cytometry data, use of viSNE (Figure 2-1) can simplify the process of distinguishing initial populations and avoid overlooking cells with unusual or unanticipated phenotypes. The second step, cell subset identification, is also traditionally performed by expert gating (Table 2-1). However, clustering tools such as SPADE (Qiu, Simonds et al. 2011) (Figure 2-2), Misty Mountain (Sugar and Sealfon 2010), and Citrus (Bruggner, Bodenmiller et al. 2014), among others, can be used to automatically assign cells to groups or clusters in high dimensional data. In the workflow here, the goal is to find all the phenotypic clusters of cells in healthy bone marrow, AML blasts, and non-blast cells from AML blood (Figure 2-2). As the final step, characterization of discovered cell subsets takes place downstream of manual gating or automated discovery tool implementation, and generally consists of feature expression comparison with heatmaps, violin plots, and histogram overlays for visualization, as well as data modeling and other statistical analysis. This workflow emphasizes integration of automated, unsupervised approaches with minimal human gating and processing. This type of semi-supervised cell population discovery and characterization can decrease human bias and variability and identify phenotypically unusual or rare cell subpopulations.

Advantages of machine learning tools: dimensionality reduction, clustering, and modeling

Not all tools perform the same analysis functions. Three functions that are useful for high-content single cell analysis include dimensionality reduction, clustering of cells into populations, and modeling. SPADE and viSNE both include dimensionality reduction steps that project multi-dimensional data into a lower dimensional space for visualization and further interpretation. These algorithms aim to preserve key high-dimensional phenotypic relationships between cells when visualizing and comparing them in 2D space. Depending on the structure of the data, other dimensionality reduction tools might be used (Table 2-1). Locally linear embedding (LLE) and isometric mapping (ISOMAP) are designed for the types of continuous phenotypic distributions seen in developmental progressions. ISOMAP accounts for geodesic distance in addition to local linear distances between high dimensional data points in order to reduce the dimensions of continuous and non-linear data (Tenenbaum, de Silva et al. 2000, Becher, Schlitzer et al. 2014). A similar principle is applied with LLE, where locally linear embedding of similar data points in high dimensional space is preserved while allowing for a non-linear global embedding of the data during projection into low dimensional space (Roweis and Saul 2000). In contrast, multidimensional scaling (MDS) and principal component analysis (PCA) preserve linear, multi-

dimensional variance. One of the advantages of PCA and other techniques, such as joint clustering and modeling (Pyne, Lee et al. 2014), is the creation of a model that can be applied to newly analyzed samples. In addition to the unsupervised tools discussed here, population analysis techniques that include some supervision can be particularly useful for mapping features across known developmental progressions (Inokuma, Maino et al. 2013, Bendall, Davis et al. 2014).

Notably, dimensionality reduction alone does not assign cells to groups. Here, dimensionality reduction with viSNE is used to aid expert interpretation of cluster identity. In this example, cells are projected onto a biaxial plot space by viSNE and then gated. Thus, viSNE is being used to see the phenotypic relationships of the cells according to all 27 protein features. This can help researchers visualize high dimensional data without losing rare populations that are best observed in single cell views. Following t-SNE or viSNE analysis, a human expert can look for cell clusters or major populations, as is the case here (Figure 2-1), or a computational tool can identify cell clusters (Table 2-1), as with t-SNE + DensVM analysis (Becher, Schlitzer et al. 2014). As the workflow becomes increasingly unsupervised, it is especially important to include a single cell view early in the analysis so that expert can perform quality checks and get a sense of the overall biological results.

Data collection, processing, and initial population identification

Data collection

In mass cytometry, as with fluorescent flow cytometry, single cell suspensions are stained with metal-conjugated antibodies specific to molecules of interest. At the mass cytometer, cells are aerosolized and streamed single-file into argon plasma where they are atomized and ionized. The resulting ion cloud passes through a quadrupole to exclude low mass ions and enrich for reporter ions whose abundance is proportional to cellular features. These reporter ions are quantified by time of flight mass spectrometry (Ornatsky, Baranov et al. 2006) and recorded in an IMD format file. These data are typically parsed into single cell events and converted to a flow cytometry standard (FCS) file for analysis (Spidlen, Moore et al. 2010). Many software programs can handle FCS files, including Cytobank (www.cytobank.org), FlowJo (www.FlowJo.com), R/Bioconductor, MATLAB, Cytoscape, and GenePattern (http://genepattern.broadinstitute.org/) (Spidlen, Barsky et al. 2013). Text files containing the expression matrix (where rows are cells and features are columns, and there is a median intensity

value for each cell) can also be extracted directly from the IMD file from the cytometer or from the FCS file for manual analysis outside of flow cytometry analysis software. In Cytobank, export of text files with intensity values is available from the FCS file details page. An expression matrix can also be extracted from the FCS file in R and MATLAB using FCS file parsing functions. In R, the package "flowCore" can be used to extract the intensity values from the FCS file using the exprs() function (Ellis B). In MATLAB, the tool "FCS data reader" includes the function fca_readfcs() to extract the intensity values of FCS files (Balkay).

Here, the healthy human bone marrow sample analyzed was obtained as a de-identified sample left over from diagnostic analysis of non-cancerous tissue in the Vanderbilt Immunopathology core. Acute myeloid leukemia peripheral blood samples were collected from consented patients. In all cases, samples are collected in accordance with the Declaration of Helsinki following protocols approved by Vanderbilt University Medical Center (VUMC) Institutional Review Board. The patient blood sample evaluated here was collected at the time of diagnosis following initial evaluation and prior to any treatment.

Data processing and scale transformation

In order to prepare data for dimensionality reduction and analysis, initial processing steps aim to ensure the quality of cell events and perform appropriate scale transformations. Quality control varies by user and is especially important when conducting studies across time or using data from different instruments. Data normalization using internal bead controls can be applied as part of this data processing (Finck, Simonds et al. 2013). In this case, the two samples were collected sequentially on the same instrument and no signal normalization was required. Efforts are underway to facilitate comparison of data among groups and centers and to report elements of panel design, instrument settings, data processing, and normalization. MIFlowCyt is a data standard set by International Society for Advancement of Cytometry (ISAC) that specifies the minimum amount of information that must be included in an FCS file to ensure reproducibility and transparency (Lee, Spidlen et al. 2008). ISAC has also established a file format for classification results from flow cytometry data (CLR) (Spidlen, Bray et al. 2015) that handles cell classification from manual or automated identification and compliments the Gating-ML file format that was developed for sharing biaxial gate classifications (Spidlen, Leif et al. 2008). Additionally, there have been efforts to standardize and compare computational flow cytometry data analysis tools. The FlowCAP project compares automated tools for cytometry data analysis using standardized

datasets (Aghaeepour, Finak et al. 2013). EuroFlow is a consortium of research groups that optimize flow cytometry protocols and analysis methods and set standards for the field of immunology and hematological studies (Kalina, Flores-Montero et al. 2012). Reporting of optimized antibody panels has also been standardized in the form of Optimized Multicolor Immunofluorescence Panels (OMIPs) (Mahnke, Chattopadhyay et al. 2010). Cytobank (www.cytobank.org) and FlowRepository (www.flowrepository.org) provide online access to annotated cytometry data files, including mass cytometry datasets (Kotecha, Krutzik et al. 2010, Spidlen, Breuer et al. 2012).

Because cytometry data are log-normal, a log or log-like scale is typically used to visualize and interpret the data. Commonly used scales include inverse hyperbolic sine (arcsinh), logarithmic, and logicle (also referred to as "bi-exponential") scales (Herzenberg, Tung et al. 2006). Logicle or log-like scales more accurately represent the spread of data around 0 than logarithmic scales, given that modern cytometers can produce negative and zero values that cannot be transformed using logarithmic scales. The implementation of the arcsinh scale here was first used for fluorescent flow cytometry (Irish, Myklebust et al. 2010) and is now standard for mass cytometry. Typically, a cofactor is included as part of the arcsinh scale transformation as a way of setting a channel specific minimum significance threshold. The cofactor is set to an intensity value below which differences are not significant. For mass cytometry, cofactors typically range from 3 to 15 and depend on background and signal to noise with the detection channel and antibody-metal conjugate. In fluorescent flow cytometry, cofactors generally range from 25 to 2000 and are especially useful in correcting for channel specific differences in spreading of negative events that depend on fluorophore selection, compensation, and instrument setup. For fluorescent flow cytometry data, appropriate compensation must also be applied prior to analysis in order to correct for any spillover between channels. Algorithms have been developed for fluorescent cytometry to automatically determine scale transformations (Parks, Roederer et al. 2004, Moore and Parks 2012). Applying an appropriate scale transformation prior to computational analysis is critical because it impacts quantification of distance between cells in the same way that it affects visualization of distance in biaxial plots.

Initial population identification and quality assessment

Beginning data analysis with a single cell view reveals the quality of the data and allows experts to spot rare cell subsets or artifacts that can be obscured in aggregate analysis. It is valuable to review the single cell data to

verify computational analysis results, and it is vital in publications to provide representative single cell views of findings. Here, intact single cells were gated by human analysis of event length and iridium intercalator uptake (Figure 2-1). This initial gating might be accomplished various ways, such as use of cisplatin exclusion to identify live cells (Fienberg, Simonds et al. 2012). Event length is generally higher for the mass cytometry equivalent of 'cell doublets' that can occur when the signal from two cells is not well separated in time. Intercalator uptake helps mark all cells and is proportional to nucleic acid content (Ornatsky, Baranov et al. 2006, Bendall, Simonds et al. 2011). A biaxial view of each channel was then used to evaluate data quality prior to computational analysis. If no intercalator positive events are seen in this view, it suggests that there were no cells in the sample or there was an error in DNA intercalator staining. Once intact, single cells have been identified (Figure 2-1A), a quick check using traditional biaxial plots or histograms can be used to ensure there is no clear overstaining. Severe overstaining results in errors while collecting data on the cytometer because event length is too great and individual cell events cannot be distinguished. Additionally, checks could be included at this step for contaminant signals. Atomic mass contaminants, such as barium and lead, can be found in water, buffers or glassware. Collecting data for the corresponding channels (137 and 138 for Ba, 208 and 209 for Pb) can be used to track these contaminants. In summary, intact single cells are first gated by a human expert. This step may be automated, but it represents an opportunity for quality assessment and initial familiarization with the data prior to computational analysis.

Unsupervised machine learning tools

viSNE

viSNE is a cytometry analysis tool that employs t-stochastic neighbor embedding (t-SNE) in mapping individual cells in a two or three-dimensional map that is based on their high dimensional relationships (van der Maaten and Hinton 2008, Amir el, Davis et al. 2013). viSNE can be used to provide a human readable two-dimensional (2D) view of cells that are arranged in a way that approximates high-dimensional phenotypic similarity. viSNE is implemented in MATLAB and Cytobank (Kotecha, Krutzik et al. 2010), and the Cytobank implementation of viSNE is shown here (Figure 2-1). viSNE can be run using a single population of cells or multiple populations drawn from one or more files. However, cell features selected for analysis must have been measured on all cell

populations in a comparable way and features must be measured on comparable scales. It is sometimes helpful to subsample cell events from populations to speed the analysis or test robustness. Sampling can be 'equal' with respect to the starting populations, in order to ensure that each cell population is represented on the viSNE map by the same number of cells, or 'proportional', so that each population is represented by a number of cells proportional to its abundance. When data are thought to contain rare cell subsets, subsampling should be avoided to preserve rare cells. Initial gating can be used to focus the analysis on a population of interest and increase its relative abundance. Here, equal numbers of cells were selected from the AML PBMC and healthy marrow files for the viSNE analysis.

The cell features selected for viSNE mapping affect the structure of the viSNE map. Markers that vary highly between cell subsets will polarize subsets, placing them farther apart in tightly grouped islands. Markers with low variance on subsets will cause those cells to be placed closer together on the map. Thus, including markers that are not expressed on any cells can result in compression of islands on the map and loss of subset polarization. Features that might contribute to clustering can be selected in an unsupervised manner based on variance. For example, features that vary more in disease than in healthy controls might be particularly useful in stratifying cells associated with distinct molecular subgroups (Irish, Hovland et al. 2004). Here, all 27 markers in the panel were included in viSNE mapping because all were expressed and variable on the cells in the samples. The displayed viSNE map shows cells from the AML patient file only (Figure 2-1). The resulting viSNE map showed a broad distribution of heterogeneous $CD45^{lo}$ AML cells and several distinct islands of non-blast cells (Figure 2-1B). Relative protein expression as heat intensity can be viewed for each marker in the panel and are shown here for the 27 markers on the panel (Figure 2-1C). The two main populations of AML blast and non-blast cells were then manually gated from the viSNE map and exported as separate FCS files for further comparison to healthy bone marrow cells using SPADE and heatmaps. All healthy marrow cells were exported from the viSNE analysis as no additional gating was required to identify major populations. Depending on the sample and biological question, it may be useful to gate initial major populations from several or all files in this step of the analysis.

The MATLAB implementation of viSNE can be accessed through the freely downloadable *cyt* tool (http://www.c2b2.columbia.edu/danapeerlab/html/cyt.html). *Cyt* employs a user interface that allows for selection

of features for mapping, selection of files or gates to be mapped, an interface for visualizing parameter intensity on a heat scale, and a tool within the interface for manually gating populations resulting viSNE map.

SPADE

SPADE is an algorithm that includes dimensionality reduction, clustering of cells into populations (also referred to as 'nodes'), and visualization using a 2D minimum spanning tree. Data must be appropriately scaled and intact cell events gated prior to SPADE analysis as described above. Here, this is done prior to viSNE gating. SPADE has been implemented in Cytobank, R, Cytoscape (http://www.cytoscape.org/), and MATLAB. In R, the package "spade" includes functions to implement individual steps of SPADE and to execute a comprehensive SPADE analysis (Linderman M). CytoSPADE is a plugin available for use in Cytoscape that provides a user interface with the R implementation (http://www.cytospade.org). The MATLAB implementation of SPADE requires the SPADE V2.0 MATLAB tool that is freely downloadable (http://pengqiu.gatech.edu/software/SPADE/index.html). Here, the Cytobank implementation of SPADE was used to compare populations identified in viSNE guided gating. User-defined parameters for SPADE analysis include downsampling, feature selection, and a target number of nodes. Target downsampling, which can be indicated as either a percentage of cells or an absolute number, specifies how much weight to give clusters of varying density. A lower downsampling percentage increases the likelihood that sparse regions of density will be given their own clusters rather than being grouped into clusters with regions of higher density. When a sample is thought to contain rare subsets of cells, entering a lower downsampling value can help distinguish these cells as a separate population (Bendall, Simonds et al. 2011, Qiu, Simonds et al. 2011). Feature selection in SPADE can also be based on selecting highly variable or biologically relevant markers, as described above for viSNE. The number of nodes indicates the target number of clusters (i.e. cell subsets) that the algorithm should produce, and 200 nodes is a good default for standard mass cytometry datasets containing ~10^5 to 10^7 total intact single cells. Including more nodes in the analysis helps to assign rare subsets to their own clusters. These clusters can be easily combined in a process called "bubbling", in which a human expert manually refines the cluster identity. A table of basic statistics, such as median intensity of each feature, is generated for each population of cells identified by SPADE and can be downloaded as a text file. Cell subsets identified by SPADE can

additionally be exported as individual FCS files for further analysis, as in the heatmap analysis shown here (Figure 2-3).

In the example here, three populations were analyzed. The two populations of AML blast and non-blast cells identified by viSNE (Figure 2-1) were compared with the population of healthy bone marrow cells stained with the same mass cytometry antibody panel. Here, a concatenated file containing all three populations was also included to allow visualization of all cells simultaneously on one tree (Figure 2-2C). SPADE can initialize with a fixed or random seed and is random in the Cytobank implementation. The same random seed can be set from run to run in the MATLAB implementation. However, when new files are added to the analysis, a different tree can still stem from the same seed, which necessitates re-running the analysis to include any additional files. For this analysis, the downsampling percentage was set to 1%, the target number of nodes was 100, and the features used for clustering were all 27 measured markers in the panel. The resulting SPADE trees are shown in Figure 2-2.

Including SPADE in this analysis workflow has several advantages. First, SPADE produces a visualization of population abundances by altering the sizes of each node depending on how many cells it encompasses. For example, it can be seen in the SPADE tree that the non-blast AML cells fall almost exclusively into one node, reflecting their relative homogeneity and the lack of normal immune cell populations in the AML patient's blood (Figure 2-2). Clustering with SPADE also assigns each cell to a discrete group, which minimizes analysis variability and prevents loss of cells that are outside of gated regions in manual biaxial gating. In a standard SPADE analysis, the algorithm is asked to "over cluster", producing hundreds of relatively small clusters rather than grouping cells into fewer, larger groups. This over clustering gives high resolution to improve rare subset identification and allows for a thorough annotation and characterization of all potentially discrete biological populations in the heatmap analysis.

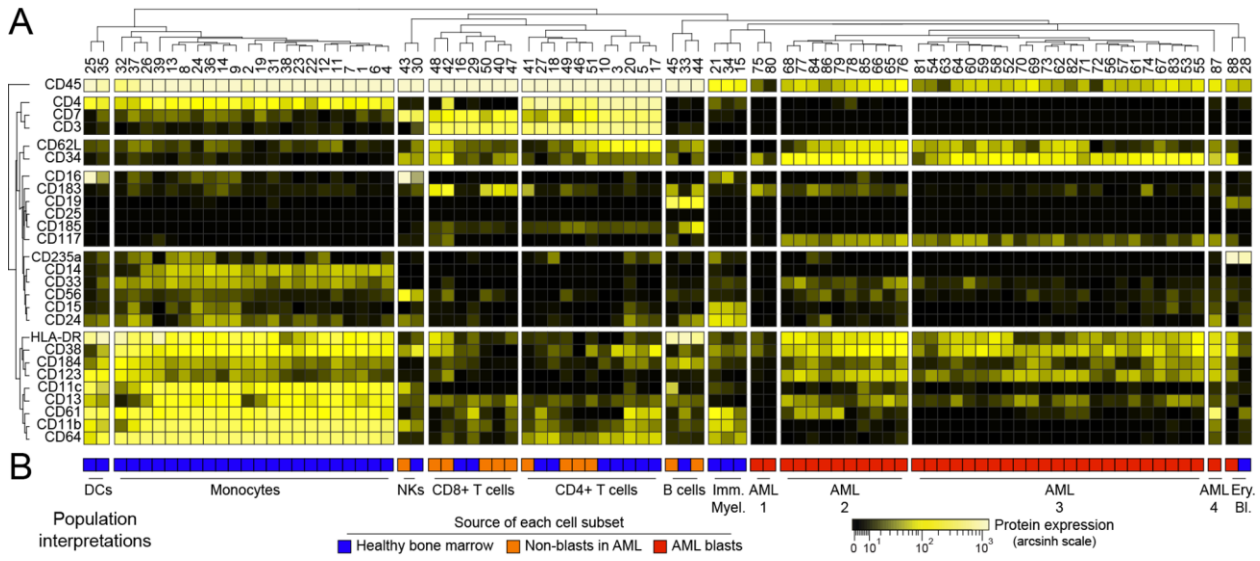

Figure 2-3. Characterizing cell subsets with a heatmap analysis of median protein expression and hierarchical clustering of proteins and populations.

A heatmap shows characterization of cell populations identified by SPADE (columns) according to median expression of 27 proteins (rows). For each sample analyzed in Fig. 2, cell populations identified by SPADE that contained at least 1% of total cells were included. Cell populations and proteins were arranged according to complete linkage hierarchical clustering. Heat intensity reflects the median expression of each protein for each cell population. B) Each population contained cells from only the indicated source (healthy marrow, non-malignant cells in AML patient blood, and AML blasts). Human interpretation of population identities based on clustered heatmap analysis is shown.

Characterizing and visualizing populations

Population heatmaps

With some algorithms it is not straightforward to compare the results of an analysis of one set of samples with the results from another set of samples. For example, with SPADE it is not straightforward to map a new sample onto an existing minimum spanning tree defined using different samples. Instead, a new SPADE analysis is generally run that includes both the new and old samples. In contrast, a heatmap can be used to compare populations identified in different analysis runs of SPADE or populations identified by different clustering techniques. Heatmaps also provide a compact view that facilitates comparing many populations according to a large variety of measured features. In heatmaps, different types of biological and clinical information can also be used to group populations or assessed for association with resulting groups (Irish, Hovland et al. 2004, Irish, Myklebust et al. 2010). While population heatmaps provide an intuitive, high-level view of the results, they can obscure variation within subsets (Kotecha, Krutzik et al. 2010). To address this, statistics other than median

expression can be shown in the heatmap, such as variance or the 95th percentile of expression (Kotecha, Flores et al. 2008, Irish and Doxie 2014).

For the last step in the workflow here, tables of statistics for the hundreds of cell subsets identified in the three starting populations (Figure 2-2) were exported from SPADE as text files listing median expression of each feature for each cell subset. Cell subsets were excluded from further analysis if they contained less than 1% of the cells in the starting population. This arbitrary threshold was set in order to exclude sparse clusters where low cell number could potentially increase the error of reported medians. Here, the 1% threshold resulted in exclusion of approximately 5% of the total cell events from heatmap characterization. The table of statistics was then imported into R using the "read.table" function from the R.utils package (Bengtsson 2015) and visualized as a hierarchically clustered heatmap using the "heatmap.2" function in the gplots package (Figure 2-3A) (Gregory R. Warnes 2015). Output of a hierarchical dendrogram as part of the heatmap can be specified as one of the input parameters of the heatmap.2 function. The R package "stats" also offers a function called "heatmap" that performs the same function as heatmap.2 with slight differences in visualization options. After the clustered heatmap was generated, expert analysis was used to assign biological classifications to each group of populations in the hierarchical clustering, and included the same populations seen in viSNE (Figure 2-1B) and SPADE (Figure 2-2B): dendritic cells (DCs), monocytes, natural killer cells (NKs), CD8+ T cells, CD4+ T cells, B cells, immature myeloid cells (Imm. myel.), four subsets of AML blast cells (AML1 through AML4), and erythroid blast cells (Ery. bl.) (Figure 2-3B).

Use of a clustered heatmap in the workflow allows for simultaneous visualization of several markers for the same clusters (population of cells) from multiple files. Furthermore, nodes are hierarchically clustered, and this clustering can be pruned at various levels by the user to further group the nodes into biological populations. It is also important to note that the distance between nodes has quantitative meaning in the clustered heatmap dendrogram, as opposed to the distances on the SPADE tree that are for visualization purposes and not quantitative. Heatmap analysis therefore compliments the SPADE visualization by facilitating simultaneous visualization of nodes from multiple files and by quantifying phenotypic distances between the nodes.

Other packages and flowCore

There are many R packages designed for statistical and visual analysis of flow cytometry data, including flowCore (Ellis B), flowViz (Ellis B), flowStats (Lo K 2009), and flowClust (Lo K 2009), among others. These packages include functions for producing heat maps, histograms, bar plots, biaxial density plots, and are part of efforts to automate and standardize computational analysis of cytometry data (Pyne, Hu et al. 2009, Aghaeepour, Finak et al. 2013). Apart from the R packages designed for flow cytometry data analysis, other analysis and visualization packages can be applied to single cell data. For example, box and whisker plots or violin plots can be produced to show median, range, and the distribution of the feature in each subset.

Other Considerations for automated flow cytometry data analysis

Algorithm Selection

Three major considerations when choosing tools or algorithms for flow cytometry data analysis include 1) linear vs. non-linear measurement, 2) supervised or unsupervised approaches, and 3) need for modeling. The first consideration is whether a linear or non-linear method of dimensionality reduction is best for the data. Phenotypic relationships between cells may follow a 'creode', or necessary path, that is non-linear with respect to protein expression (i.e. co-expression or co-variance of molecules is not linearly correlated with important progressions in cellular identity or trajectories in data space) (Irish 2014). In this case, nonlinear dimensionality reduction tools may better preserve the high dimensional phenotypic relationships between cells compared to tools that assume a linear relationship between variables. The second consideration is whether an unsupervised or supervised method is needed. In an exploratory analysis where novel populations are anticipated, unsupervised approaches will minimizes the risk of overlooking the populations. Lastly, a consideration is whether or not the goal of analysis is to build a model. Mixture modeling tools can be implemented for analysis of flow cytometry data that will produce a model as output for downstream analysis. Additional issues to consider include 1) selection of features, which is generally initiated by hypotheses and pragmatic concerns and then narrowed to include those features with biologically meaningful variation (Irish, Hovland et al. 2004), and 2) aspects of statistical power, including sample size, cluster density, and false discovery rate (FDR). It is vital to calculate FDR or a related statistic, such as the f-measure, in cases where a truth is known (Aghaeepour, Finak et al. 2013).

Scalability of Workflow

Biomedical studies that employ flow and mass cytometry often accrue large numbers of samples over long periods of time. This and similar workflows can be adapted to accommodate data from these large studies. In order to account for experimental or instrument variability, normalization is necessary in these cases in order for compare samples run at different times or from different instruments. Bead normalization has been optimized for use with mass cytometry to control for machine variability between runs (Abdelrahman, Dai et al. 2009, Abdelrahman, Ornatsky et al. 2010, Finck, Simonds et al. 2013). Polystyrene beads embedded with heavy metal isotopes are run with every sample as a standard that can be used to correct MI values for each event based on technical variability. When samples accrue over a long period of time, a key consideration is that new results may not be easily mapped back to the original viSNE map or SPADE tree without re-analysis. This is one advantage of heatmaps, which compare samples according to a simple 'model' of the data, such as median expression of selected features.

This workflow as presented includes manual intervention that could be prohibitive when analyzing many data files simultaneously. While all steps of this analysis could generally be batched and automated, human review of single cell data is advantageous at workflow breakpoints to verify computational results and spot artifacts. Cytobank and other flow cytometry data analysis software allow for rapid, simultaneous viewing and pre-processing of multiple files, including scale transformation and gating. viSNE analysis can currently be run on up to 800,000 cells in Cytobank, and this limit is pragmatic, not theoretical. Many files can be run simultaneously by subsampling cells equally or proportionally from each file prior to the viSNE run. SPADE can also be run on many files simultaneously, and data files with cluster information can be quickly downloaded in a compressed folder.

Import of text files into R and selection of nodes based on the number of cells they contain can be automated and batched for highly scalable and rapid heatmap analysis. However, a potential limitation of large-scale analyses is the visualization of all nodes simultaneously on the heatmap. It may be useful in these cases to segment the SPADE tree into major populations by "bubbling" and then building separate heatmaps from each bubble rather than for the whole tree. Depending on the expected prevalence of rare cells in the dataset, the

user can request fewer nodes in the SPADE run in order to decrease the final number of clusters to be analyzed and visualized on the heatmap.

Conclusions

Data analysis in cytometry remains largely manual, supervised, and focused on large changes in magnitude of expression. As new tools are developed to assist in gating, reduce dimensionality, and automate analysis, it is important to show biologists the value of these tools and to integrate them into workflows that can become routine. The workflow presented here blends supervised and unsupervised analysis tools so that biologists can visualize results at the single cell level while still getting an accurate view of the big picture. Combining tools also allows the analyst to visualize data in multiple ways, which can be useful to extract the most meaning from a data set. Existing tools allow for identification of populations based on single cell expression profiles and characterization of these subsets using standard statistics, including expression magnitude, marker variance, and subset abundance. Going forward, tools that quantify cellular heterogeneity, identify critical population features, and assign biological identity to machine-identified subsets will be particularly useful in filling out the toolkit.

Acknowledgements

This study was supported by R25 CA136440-04 (K.E.D.), NIH/NCI K12 CA090625 (P.B.F), R00 CA143231-03 (J.M.I.), the Vanderbilt-Ingram Cancer Center (VICC, P30 CA68485), and VICC Young Ambassadors and VICC Hematology Helping Hands awards. Thanks to Mikael Roussel for helpful discussions of myeloid cell identity markers.

Conflicts of Interest: JMI has a financial interest as co-founder and board member in Cytobank Inc., a software company for single cell data analysis. No other conflicts.

CHAPTER 3

HIGH-DIMENSIONAL ANALYSIS OF ACUTE MYELOID LEUKEMIA REVEALS PHENOTYPIC CHANGES IN PERSISTENT CELLS DURING INDUCTION THERAPY

Authors: P. Brent Ferrell, Jr., Kirsten E. Diggins, Hannah G. Polikowsky, Sanjay R. Mohan, Adam C. Seegmiller, and Jonathan M. Irish

This work appears as published in manuscript form in *PloS ONE* 2016 (open access).

http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0153207

Preface

Increased heterogeneity within tumors has been found to correlate with poor therapy response and poor outcome across a variety of cancers. However, these studies primarily focused on genomic alterations in aggregate tumor samples, and genetic mutations or other alterations do not always linearly correlate with altered protein expression levels. It would therefore be useful to develop metrics of cellular heterogeneity based on single-cell proteomic data, such as that generated from flow cytometry analysis. In order to address this need, I worked closely with Dr. Brent Ferrell, then a medical fellow in Jonathan Irish's lab at Vanderbilt University, on a project tracking phenotypic heterogeneity in acute myeloid leukemia (AML) blast cells over the course of induction therapy in multiple AML patients. Dr. Ferrell aimed to determine whether changes in phenotypic heterogeneity correlated with patient therapy response, and I chose to collaborate with him on this project because it provided an ideal use case for the development and application of a single-cell level metric of sample heterogeneity. Together, we developed the conceptual framework for quantifying cellular heterogeneity based on distances between cell populations in reduced-dimensional phenotypic space. I developed a workflow for this analysis using viSNE in Cytobank (www.cytobank.org) as the first step and, for the second step, an R script that calculated the Euclidean distances between pre-defined population centers on viSNE maps. Dr. Ferrell and I applied this workflow and R script to the data presented here, and we ultimately found that an increased phenotypic distance between blast cells and normal hematopoietic stem cells correlated with poor patient outcome.

Abstract

The plasticity of AML drives poor clinical outcomes and confounds its longitudinal detection. However, the immediate impact of treatment on the leukemic and non-leukemic cells of the bone marrow and blood remains relatively understudied. Here, we conducted a pilot study of high dimensional longitudinal monitoring of immunophenotype in AML. To characterize changes in cell phenotype before, during, and immediately after induction treatment, we developed a 27-antibody panel for mass cytometry focused on surface diagnostic markers and applied it to 46 samples of blood or bone marrow tissue collected over time from 5 AML patients. Central goals were to determine whether changes in AML phenotype would be captured effectively by cytomic tools and to implement methods for describing the evolving phenotypes of AML cell subsets. Mass cytometry data were analyzed using established computational techniques. Within this pilot study, longitudinal immune monitoring with mass cytometry revealed fundamental changes in leukemia phenotypes that occurred over time during and after induction in the refractory disease setting. Persisting AML blasts became more phenotypically distinct from stem and progenitor cells due to expression of novel marker patterns that differed from pre-treatment AML cells and from all cell types observed in healthy bone marrow. This pilot study of single cell immune monitoring in AML represents a powerful tool for precision characterization and targeting of resistant disease.

Introduction

Acute myeloid leukemia is one of the deadliest adult cancers. The five-year overall survival is 21.3% for all ages and 4.6% for individuals 65 and older (2013). Current standard of care therapy has remained relatively unchanged over the last 30 years despite efforts to improve these poor outcomes (Stein and Tallman 2012). AML genetic heterogeneity has been well characterized as contributing to poor outcomes (Valk, Verhaak et al. 2004, Patel, Gönen et al. 2012, Walter, Shen et al. 2012), and longitudinal genetic analyses have suggested multiple models of clonal evolution to explain disease aggressiveness (Ding, Ley et al. 2012, Klco, Spencer et al. 2014). While it is clear that cell subsets within a pretreatment leukemia cell population have differential responses to therapy, it is not known to what extent genetic and non-genetic cellular features confer these differential responses. A single-cell understanding of AML therapy response over time during early treatment will

characterize how different treatments reprogram AML cell phenotypes and impact clonal dynamics. Immediate post-treatment changes may have lasting impacts on long term outcomes, and a better understanding of how AML cells change following treatment may highlight key targets of opportunity for new treatments. Mass cytometry and unsupervised tools from machine learning provide a new opportunity to comprehensively characterize cellular heterogeneity and improve our understanding of how different treatments impact AML cell biology. In particular, it would be useful to characterize AML cells that remain immediately following treatment and determine whether they are distinct in a way that might be therapeutically targeted.

Immunophenotype characterization by flow cytometry has become part of standard of care in AML for diagnosis and disease monitoring, and standard antibody panels have been published for traditional fluorescence flow cytometry used in clinical pathology (Stetler-Stevenson, Davis et al. 2007, van Dongen, Lhermitte et al. 2012). A key strength of flow cytometry is the ability to measure several independent properties on each cell and to use complex combinations of these quantitative measurements to classify or isolate cells of interest (Krutzik, Irish et al. 2004, Irish, Kotecha et al. 2006). Surface antigens, such as CD34 and CD123, have been extensively studied individually or in small combinations, but reported associations with clinical outcome are numerous and often conflicting (Mason, Juneja et al. 2006). Additionally, leukemia stem cells (LSCs) and stem-ness properties likely play a significant role in therapy resistance and leukemia persistence in AML (Eppert, Takenaka et al. 2011). Furthermore, the markers that characterize an AML at diagnosis may shift during treatment and be changed dramatically in the case of minimal residual disease (MRD) or relapse (Baer, Stewart et al. 2001). Small antibody panels focused on positive identification of AML cells are susceptible to overlooking AML clones that undergo antigenic changes.

In contrast, measurement of 30 or more features by mass cytometry (Bendall, Simonds et al. 2011) can comprehensively characterize normal myeloid cell populations and, in combination with unsupervised machine learning tools, robustly characterizes all non-AML cells and distinguishes them from AML blasts (Becher, Schlitzer et al. 2014). Mass cytometry thus gives the potential for improved ability to define subsets throughout therapy. Because high-dimensional mass cytometry generates significantly more data than a traditional flow cytometry experiment, it therefore creates the need for new data processing and visualization tools. Computational tools, especially unsupervised algorithms, organize and display high dimensional data in a way not possible with

traditional, supervised gating techniques (Qiu, Simonds et al. 2011, Amir el, Davis et al. 2013, Bruggner, Bodenmiller et al. 2014, Shekhar, Brodin et al. 2014). One such algorithm, viSNE, has been shown to be robust in its ability to distinguish both healthy and leukemia subsets, showing great promise for research and clinical analysis and visualization of cytometry data (Amir el, Davis et al. 2013, McCarthy 2013). viSNE creates a phenotypic map of cells from an individual sample, or collection of samples, enabling visualization of phenotypic relationship between individual cells (Amir el, Davis et al. 2013). Furthermore, we have validated and published methods to use this tool to analyze both healthy and leukemia samples (Diggins, Ferrell et al. 2015). Given viSNE's particular strength in visualization of high dimensional single cell data, it is well-suited to identify subtle or large changes in marker expression across several samples. Additionally, cells with unexpected phenotypes are routinely overlooked in manual analysis and viSNE captures many of these overlooked cells (Amir el, Davis et al. 2013, Irish 2014).

We present a comprehensive single cell view of AML that tracks changes in the bone marrow and blood over time before, during, and after treatment and relates differences in cellular phenotype between patients over time and between cells from AML patients and healthy donors. From each patient, up to 12 samples of blood and marrow were obtained before, during, and after induction chemotherapy (n = 46 total samples; 14 marrow and 32 blood from 5 individual patients). This analysis represents a pilot of in-depth longitudinal monitoring of AML during induction therapy with mass cytometry. Additionally, machine learning analysis tools and mass cytometry were used to create a phenotypic stem-ness index for AML and applied to evaluate longitudinally collected patient samples.

Materials & Methods

Patients and Consent

Patients with suspected AML were consented for protocol sample collection. Healthy bone marrow was obtained from leftover sample from diagnostic analysis of non-cancerous tissue in the Vanderbilt Immunopathology core. All specimens were obtained in accordance with the Declaration of Helsinki following protocols approved by Vanderbilt University Medical Center (VUMC) Institutional Review Board (IRB). For each patient participating in the collection protocol, signed, written consent was obtained via a written consent form that was also approved by the VUMC IRB. Eligibility criteria included >=18 years of age with suspected acute myeloid leukemia

undergoing clinical evaluation at VUMC. Patients diagnosed with AML, excluding acute promyelocytic leukemia, who were treated with intensive induction chemotherapy, were eligible for complete collection protocol, regardless of induction regimen (individual regimens noted in Table 3-1). Basic clinical characteristics for each patient are listed in Table 3-1. One normal bone marrow was obtained from Vanderbilt Hematopathology Lab.

Table 3-1 – Patient characteristics and clinical outcomes

ID	Age	Sex	Alive§	Survival (months)	BM Blast	WBC	PB Blast	Cytogenetics	FLT3	NPM1	Induction 1	Induction 2	Relapse
F001	23	F	+	14.5	97.5	432 *	95	46,XX,t(9;11)(p22;q23)[20]	Neg	Neg	CR	N/A	N/A
F002	77	M		9	26	3.2	23	42~47,X,-Y,add(2)(p11.2),-5,add(6)(p12),add(7)(q22),add(12)(p13),del(12)(p12),-15,-18,-19,-20,+1~5mar	Neg	Neg	CR	N/A	3
F003	61	M		4.5	88.5	15.7	94	46,XY,del(11)(q21q23)[2]/46,XY[20]	Neg	Neg	Residual Leukemia	Residual Leukemia	N/A
F004	49	F	+	15	45	5.7	28	46, XX[20]	Neg	Neg	Residual Leukemia	CR	N/A
F007	59	F	+	14	52	1.8	0	47,XX,+22[3]/46,XX[17]	Neg	Neg	CR	N/A	N/A

§ '+' indicates patient was alive at time of last follow-up.

* Patient F001 underwent leukapheresis prior to treatment and sample collection.

After patients were consented, bone marrow and peripheral blood samples were obtained per protocol. Peripheral blood samples were collected at time of diagnosis or initial evaluation, every 2-3 days during the first two weeks of induction, day 14 (mid-induction) and at time of hematologic recovery, if applicable (Figure 3-1A). Bone marrow (BM) was collected at all clinically indicated time points, including diagnosis, mid-induction (day 14), and hematologic recovery, if applicable.

Sample Processing and Preservation

Once obtained, samples underwent immediate density gradient separation of mononuclear cells using BD Vacutainer® CPT™ Cell Preparation Tube with Sodium Heparin (BD Biosciences, Franklin Lakes, NJ). The separated mononuclear cells were then pelleted with low speed centrifugation (200 x g) and aliquoted into

multiple cryotubes in an 88% fetal bovine serum (FBS) + 12% DMSO solution. Samples were stored at -80 C for 24-72 hours prior to long-term storage in liquid nitrogen.

Mass Cytometry

Aliquots of cryopreserved samples were thawed into 10ml of warm media (90% RPMI 1640 (Mediatech, Inc., Manassas, VA) + 10% fetal bovine serum (Gibco® standard FBS, life technologies, Grand Island, NY)), pelleted by centrifugation at 200 x *g*, washed with warm media and pelleted again at 200 x *g* before resuspension in flow cytometry tubes with warm media (Falcon 2052, BD-Biosciences, San Jose, CA) and allowed to rest for 30 minutes in a 5% CO_2 incubator at 37 °C. Each rested sample was then pelleted at 200xg, washed with phosphate buffered saline (HyClone®, HyClone Laboratories, Logan, UT), pelleted and resuspended in cell staining media (CSM=PBS + 1% bovine serum albumin (BSA)) (Fisher Scientific, Fair Lawn, NJ). Cells were stained with a mass cytometry antibody panel of 27 antibodies (DVS Sciences, Sunnyvale, CA) designed based on inclusion of both consensus standard of care immunophenotyping panels for AML, as well as antibodies that allowed identification of non-AML PBMC (Table 3-2) (Stetler-Stevenson, Davis et al. 2007). A master mix of these antibodies was added to each sample to give a final staining volume of 50µL and incubated at room temperature for 30 minutes. Cells were then washed twice, first with CSM and then with PBS and then permeabilized in -20 °C 100% methanol for 20 minutes. Following permeabilization, cells were washed, stained with 250 nM Iridium intercalator (ORNATSKY, LOU ET AL. 2008) (Fluidigm, San Francisco, CA) for 16 hours at 4 °C, washed twice in PBS, and then re-suspended in 500 µL ddH_2O for CyTOF analysis. Samples were analyzed using a CyTOF 1.0 at the Vanderbilt Flow Cytometry Shared Resource.

Table 3-2 –
AML mass cytometry panel

Target (Clone)	Type	Mass
CD235a (HIR2)	β	141
CD19 (HIB19)	γ	142
CD117 (104D2)	α	143
CD11b (ICRF44)	α	144
CD4 (RPAT4)	β	145
CD64 (10.1)	β	146
CD7 (CD7-6B7)	α	147
CD34 (581)	α	148
CD61 (VI-PL2)	β	150
CD123 (6H6)	γ	151
CD13 (WM15)	α	152
CD62L (DREG-56)	γ	153
CD45 (HI30)	α	154
CD183 (G025H7)	γ	156
CD33 (WM53)	α	158
CD11c (Bu15)	γ	159
CD14 (M5E2)	α	160
CD15 (W6D3)	α	164
CD16 (3G8)	α	165
CD24 (ML5)	γ	166
CD38 (HIT2)	α	167
CD25 (2A3)	β	169
CD3 (UCHT1)	γ	170
CD185 (51505)	α	171
HLA-DR (L243)	α	174
CD184 (12G5)	γ	175
CD56 (CMSSB)	α	176
Iridium	§	191

Marker type key:
α – Consensus AML
β – Secondary AML
γ – Other
§ - Mass cytometry

Table 3-2. Table of antibodies for mass cytometry**.** Antibodies used for mass cytometry staining are listed in this table. Clone of the antibody along with designation of type and mass number are listed for each antibody. Antibody "type" in this table is based on markers recommendations for AML phenotyping by the Bethesda International Consensus Conference (Wood, Arroz et al. 2007).

Data Analysis

Mass cytometry data (.fcs) files were evaluated using R (version 2.5.2, The R Foundation for Statistical Computing), viSNE in MATLAB (version R2013b, The Mathworks, Inc.), and Cytobank software (Cytobank Inc.) (KOTECHA, KRUTZIK ET AL. 2010, AMIR EL, DAVIS ET AL. 2013). Files from each patient were mapped using the viSNE MATLAB graphical user interface, *cyt*, available at http://www.c2b2.columbia.edu/danapeerlab/html/cyt.html. Files used for viSNE co-analysis had the same number of cells randomly sampled from each file, unless very few cells were collected for a particular time point, in which case all cells were used in viSNE analysis.

Measuring Hematopoietic Stem Cell Distance within viSNE

Within a given viSNE map, the hematopoietic stem cell distance (HSCD) for a given cell was calculated as the standard Euclidean distance formula ($\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$) where x_1, y_1 was the coordinate location of a cell on the viSNE map's t-SNE axes and x_2, y_2 was the center of the healthy HSC population on the same viSNE map. In particular, $CD34^{+}$ $CD38^{lo/-}$ hematopoietic stem cells were used as a common reference point for comparing non-leukemia subsets and phenotypically diverse AML samples. The center of the healthy HSC population was determined by averaging the x and y coordinates of the HSCs in the viSNE map. A population's HSC distance was calculated as the mean HSC distance of all cells in that group. For comparison of healthy and leukemia inter-sample heterogeneity, mass cytometry .fcs files of healthy PBMC (Nicholas, Greenplate et al. 2015) and bone marrow (Amir el, Davis et al. 2013) were used from published sources and remapped in separate viSNE analyses. Distances of common cell populations were measured from designated cells populations, HSC in the case of marrow samples and CD4+ T cells in the case of PBMC. Distances of these populations found in each sample were measured from reference population as described above and median and interquartile range (IQR) values were calculated to give a measure of inter-sample heterogeneity within healthy and AML samples.

Results

Machine Learning and Mass Cytometry Separated AML Blast Cells from Non-Malignant Cells

AML blast cells were identified in viSNE using a combination of marker expression and localization on the 2–dimensional viSNE map. In each sample, a discrete group of cells was identified in the viSNE map characterized by a phenotype consistent with myeloid blasts (Figure 3-1B). Measured blast percentages were consistent with those reported for clinical pathology evaluation (data shown for F001 and F003 in Figure 3-5). Furthermore, other discrete islands within the map represented known populations of non-malignant mononuclear cells if present (Figure 3-1B and 3-1C). These islands were high for CD45 expression. Each island was characterized by analysis of each marker measured in that population and was assigned a cell identity (Figure 3-1B).

viSNE Distance from Stem Cells Provides a Common Measure of Phenotypic Similarity to Stem Cells

In a viSNE map, proximity corresponds to phenotypic similarity according to the 27 measured proteins selected when creating the viSNE axes. To investigate the phenotypic similarity of non-leukemia mononuclear cells to stem cells, healthy bone marrow mononuclear cells and gated blast cells from four diagnostic AML bone marrow samples were mapped with viSNE (Figure 3-2A). Populations identified within a healthy sample included hematopoietic stem cells (HSCs), early progenitors (EPs), dendritic cells, maturing myeloid cells, monocytes, natural killer (NK) cells, CD4+ and CD4- T cells and B cells. Phenotypic distance from HSCs was then used to characterize healthy and AML population. Healthy NK cells and T cells were the most phenotypically distinct from HSC with distances of 46.6 (NK), 54.1 ($CD4^+$ T cells), and 60.4 ($CD4^-$ T cells). $CD34^+$ $CD38^+$ bone marrow cells comprised the majority of the EP population. EPs occupied the largest area of phenotypic space relative to the number of cells in the population, indicating that EPs were the most heterogeneous cell population observed. The mean distance of EPs from HSCs was 12.4, indicating that, on average, EPs were the healthy cell type most closely related to the HSCs, as expected. EPs were phenotypically similar to healthy stem cells when considering 27 protein features (Figure 3-2B). This calculated HSC distance was used to quantify the phenotypic distance of each patient's AML blasts from known cell populations within healthy marrow.

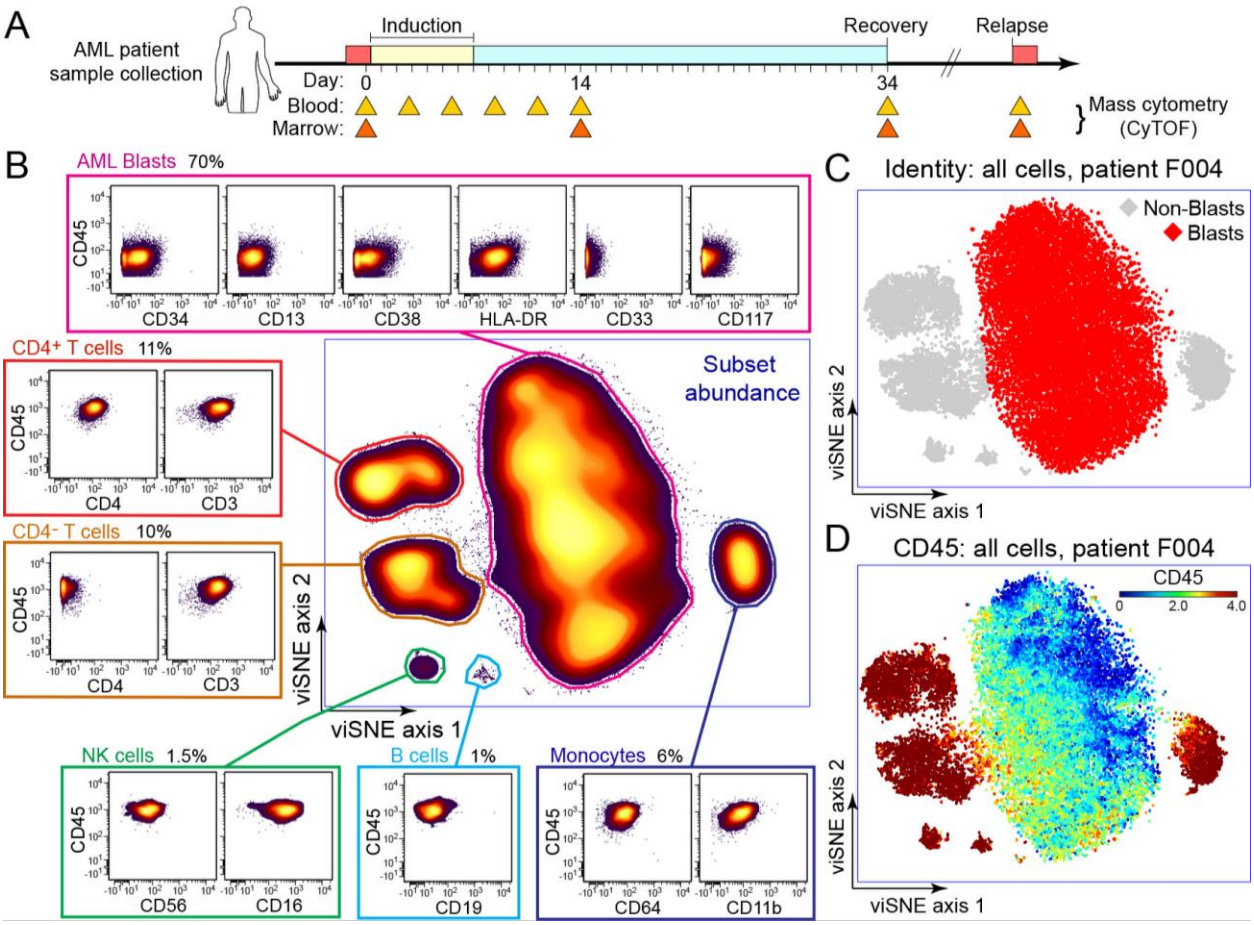

Figure 3-1. Overview of mass cytometry phenotyping in early AML therapy

(A) A timeline of AML induction shows blood and bone marrow collection scheme for this study. (B) viSNE analysis of all live cells from the diagnosis marrow of one AML patient F004 is shown. Cells were arranged on the viSNE map along unitless x and y viSNE axes according to 27-dimensional phenotype (Table 3-2) so that phenotypically similar cells were placed near each other. Cellular abundance is indicated with a shaded contour plot where outliers start at 10% and each 2% contour is shaded a lighter color from purple to yellow. (C) On the same viSNE axes as in (B), diagnostic bone marrow cells from patient F004 were graphed and shaded according to identity determined by immunophenotype. AML blast cells were shaded red and non-blast cells were shaded grey. (D) On the same viSNE axes as in (B), diagnostic bone marrow cells from patient F004 were graphed and shaded according to expression of CD45 on a rainbow heatmap (log-like arcsinh_{15} scale).

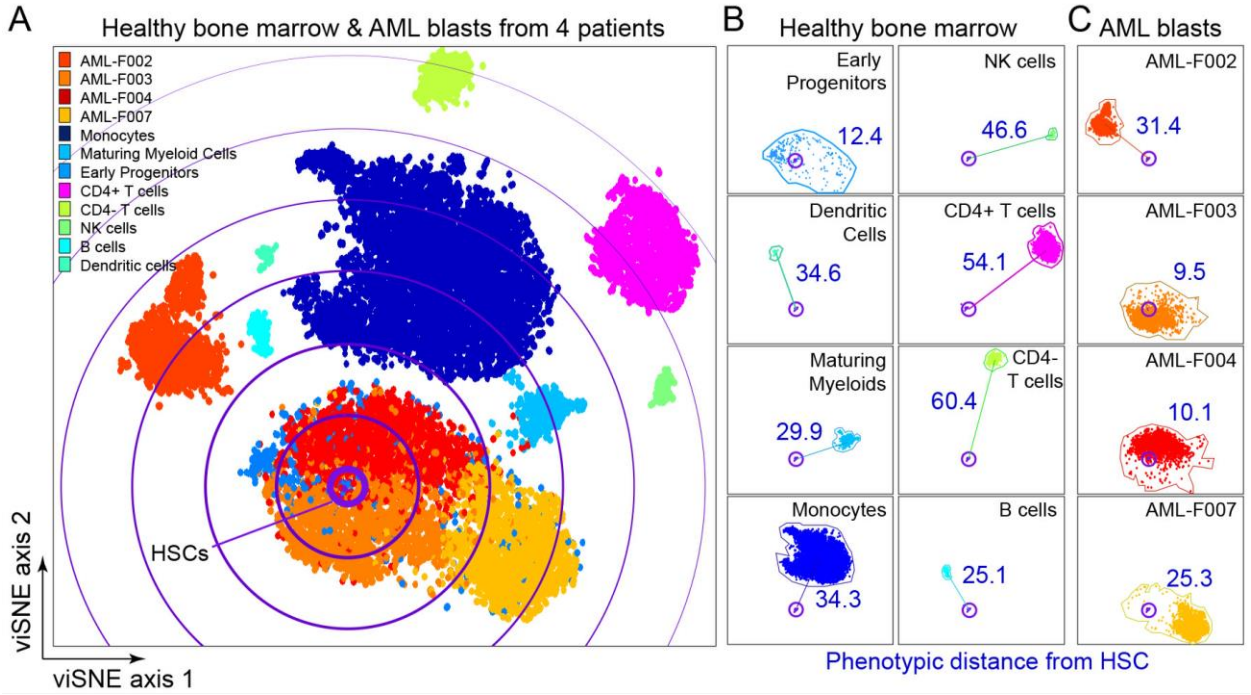

Figure 3-2. Phenotypic distance from hematopoietic stem cells distinguishes healthy cell populations and AML blasts from different individuals. (A) A 27-dimensional viSNE analysis compares an equivalent number of live cells from normal bone marrow and from each of four AML patient bone marrow samples obtained at diagnosis prior to treatment. (B) 27-dimensional phenotypic distance of normal bone marrow mononuclear populations from healthy hematopoietic stem cells (HSCs) was measured in the viSNE analysis from (A) and is shown in blue. (C) As in (B), the HSC distance for the blast populations from four AML patients was measured and is shown in blue.

Diagnostic AML Blasts from Four Patients Differed in Phenotypic Distance from HSCs

Within the viSNE map, HSC distance was measured for AML blasts from four patients' diagnostic samples (Figure 3-2C). AML blasts localized to one general area of the viSNE map; however, while samples from F003, F004, F007 were associated closely, AML blasts from patient F002 were phenotypically distinct and well separated from other AMLs (Figure 3-2A). In contrast with healthy early progenitor cells, AML blasts were more homogeneous in phenotype, as AML blasts occupied a smaller phenotypic area of the map than EPs (Figure 3-2A). This was despite the fact that AML blast populations contained approximately 5.4 times as many cells as the EP population in this analysis.

AML blasts from patients F003 and F004 were phenotypically similar to healthy HSCs and had very low distance values near 10. This result contrasted with that from AML patients F002 and F007, whose diagnostic blasts were farther from healthy stem cells and had distance values near 30 (Figure 3-2), highlighting the patient-to-patient diversity that can be detected with even a small cohort of patients. By comparing distance to HSC and

other populations on the viSNE map, each patient's AML blasts were characterized as phenotypically more similar to HSCs (F003, F004) or more distant from HSCs (F002 and F007) (Figure 3-2B).

To assess donor variation for this type of comparison for other cell subsets, distances on viSNE axes were compared for 5 sets of test and reference populations from this manuscript and others and reported as the interquartile range (IQR) and median across individuals. Little heterogeneity in viSNE distances was observed when comparing CD4+ T cells from healthy PBMC to B cells from healthy PBMC (IQR: 1.3, median: 38.1, N = 7). Similarly, relatively little heterogeneity was seen in comparisons of healthy BM B cells to healthy BM HSCs (IQR: 4.5, median: 10.1, N = 3), non-leukemia CD4+ T cells from AML BM to healthy BM HSCs (IQR: 2.7, median: 9.5, N = 3), and non-leukemia CD4+ T cells from AML BM to healthy BM HSCs (IQR: 2.4, median: 7.3, N = 3). All of these low interquartile ranges contrasted with the significantly greater heterogeneity seen when comparing AML blasts compared to healthy BM HSCs here (IQR: 18.6, median: 17.7, N = 4, Figure 3-2). Thus, the heterogeneity in viSNE distance was specific to the leukemic blasts and not observed in comparisons including non-leukemia cells from the same sample, healthy blood, or healthy bone marrow.

Mapping Early Treatment Samples in viSNE Allowed Accurate Assessment of Blast Clearance

In order to assess whether this technique properly characterized blast clearance, samples over time from individual patients were mapped in patient-specific viSNE analyses. A leukemic blast area or island was identified in the viSNE map and treatment response was observed as the regression or persistence of cells within this area over time. In patients who achieved remission (based on clinical pathology assessment), >95% of cells in the recovery samples were outside of the leukemic blast area (Figure 3-3, Figure 3-4, Figure 3-5).

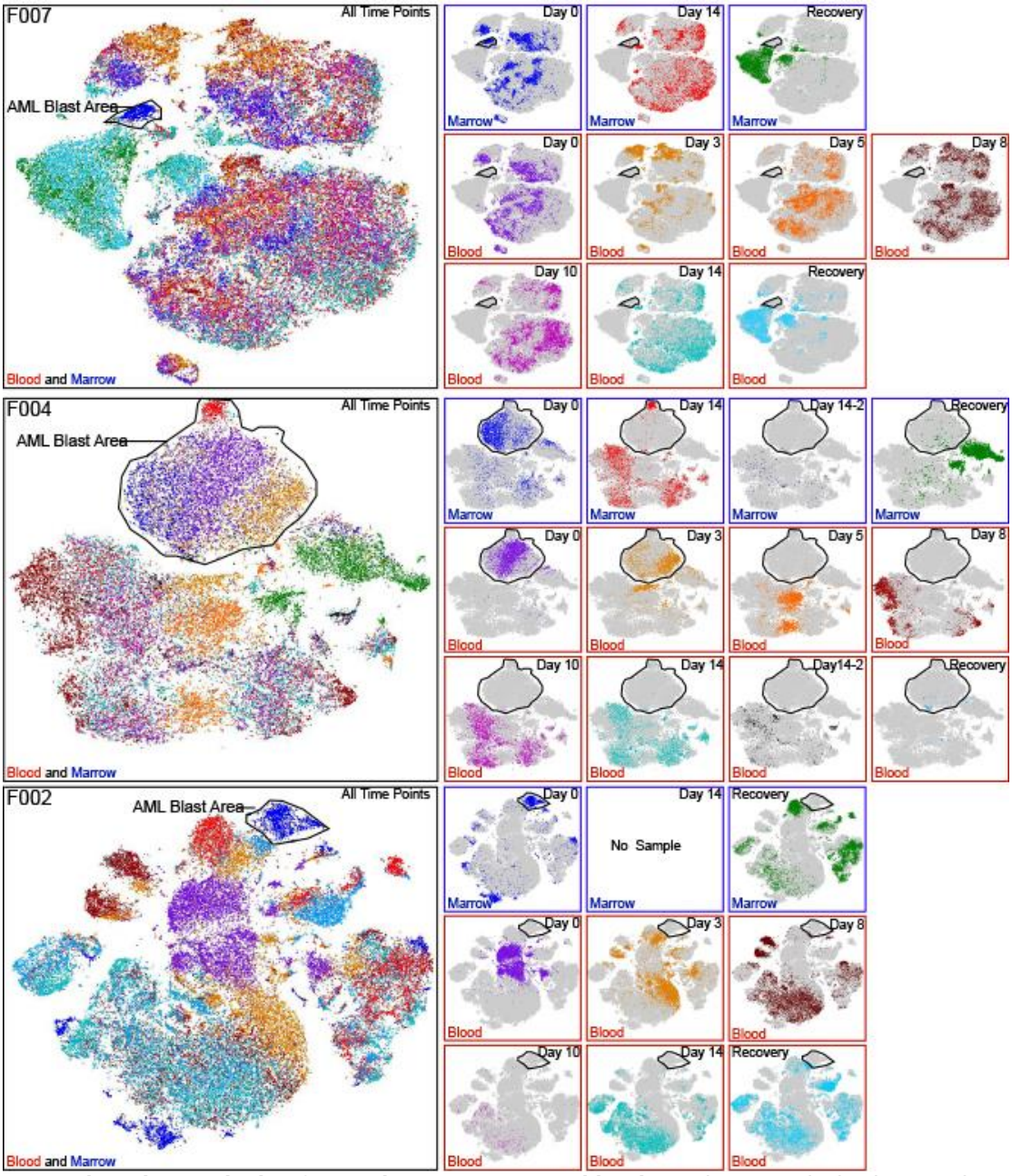

Figure 3-3 – Comprehensive analysis of AML therapy response kinetics. viSNE analysis characterizes changes in leukemia cell phenotype over time for cells from all treatment times for three individuals. All cells from all clinical timepoints were analyzed using viSNE according to the 27 markers measured (Table 3-2). An AML blast area was identified as in Figure 3-1. Color indicates clinical time point and source, either bone marrow or blood. Patients F002 and F007 had very low blast percentages in the peripheral blood at diagnosis.

Abundance and phenotype of AML cells changed significantly in the blood during the first days of treatment, with significant variability among patients (Figures 3-3 & 3-4). Absolute numbers of all cell types, malignant and non-malignant, declined predictably in the blood of all patients by Day 14. Lymphocytes, predominantly T cells, were the most abundant cell population in all patients at Day 14, when leukocyte counts in marrow and blood are

typically at their nadir. Leukemic blasts in the blood and marrow formed distinct islands in viSNE and these islands could be followed over time for each patient ("Leukemic Blast Area", Figures 3-3 & 3-4). Cells were observed largely to regress from this area in blood by day 14 in all patients (>90%). The patient with the highest Day 14 blast population (F003) also had the highest percentage of blasts seen in the bone marrow (see below). Overall, the mass cytometry results were comparable to those obtained by clinical pathology analysis (Figure 3-5).

Mid-induction (Day 14) bone marrow aspirate was obtained from patients when sufficient sample was available (patient F002 had inaspirable marrow at day 14, though a core biopsy revealed no residual AML). Bone marrow from patients F003 and F004 was significantly involved with AML post-induction, and both patients required reinduction. Patient F003 never achieved remission despite two inductions and the initial Day 14 marrow was highly involved by AML (90% blasts by microscopy, 81% by clinical flow cytometry). This clinical finding was corroborated by mass cytometry and viSNE analysis, as both patients had cells mapped within the leukemic blast area at Day 14 (Figures 3-3 & 3-4). For patient F004, residual cells occupied a tight area on the map and were phenotypically distinct from the majority of the AML cells in the diagnostic sample, both in the marrow and the blood (Figure 3-3). This residual population of AML cells did not persist in the second mid-induction biopsy (Day 14-2) or at recovery biopsy (after second induction), when the (Amir el, Davis et al. 2013)(Amir el, Davis et al. 2013)(Amir el, Davis et al. 2013)(Amir el, Davis et al. 2013)(Amir el, Davis et al. 2013)(Amir el, Davis et al. 2013)(Amir el, Davis et al. 2013)patient was found to be in remission. Thus, a population of leukemia cells from F003 and F004 persisted after the initial treatment and a change in overall AML cell phenotype was observed in this persistent population in both cases.

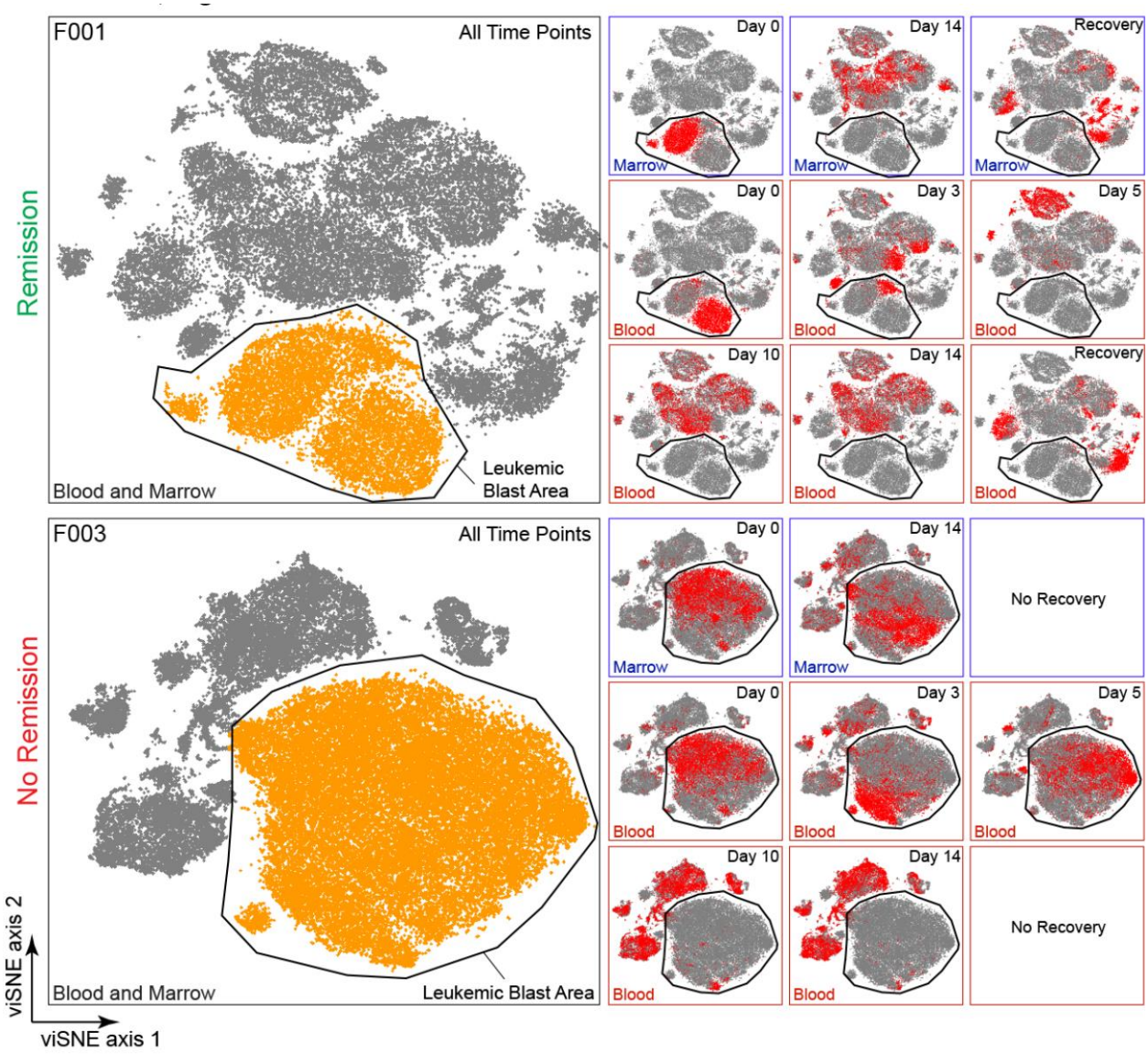

Figure 3-4. Computational analysis of samples throughout induction allows visualization of both remission and persistent AML. 27-dimensional viSNE analysis of all live cells from all sample collection time points for two AML patients is shown (left). Clinical response is indicated for each patient on the left. Leukemic blast areas were determined by location of cells at diagnosis and analysis of marker expression, as in Figure 3-1. At right, cells taken from the time point shown (red) were compared to all cells (grey). Differences in the location of cells within the viSNE map resulted from changes in protein expression.

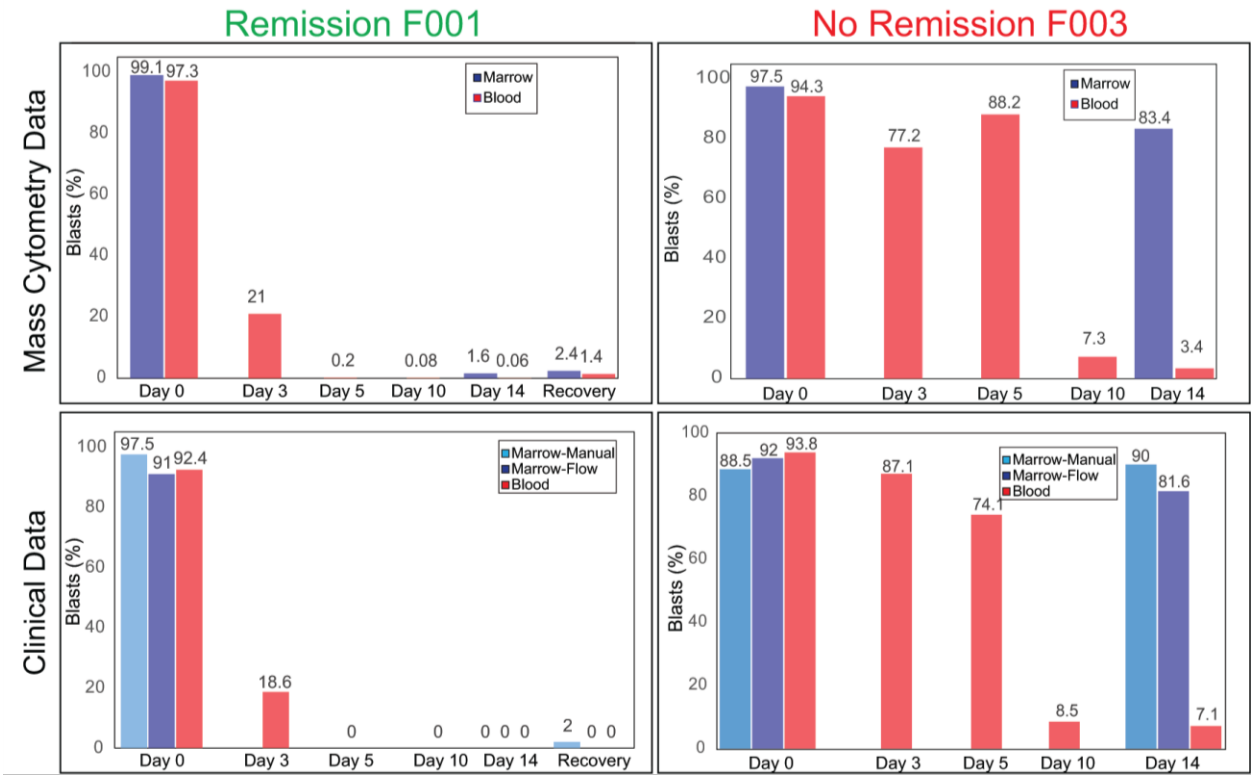

Figure 3-5. Analysis of immunophenotypic change throughout induction - F001 and F003. At top, mass cytometry quantification of blast percentage for each sample from two patients is shown. Below, clinical cytometry and microscopy data of each blood (clinical samples collected at different time on same day) and bone marrow (clinical and research samples collected at same time) sample throughout induction is shown. The patient whose samples are displayed on the left (F001) achieved remission (defined as <5% marrow blasts at recovery). On the right, a refractory patient's samples are shown, in whom a very high blast percent were seen at day 14.

Mass Cytometry Identified Phenotypically Distinct Persisting AML Blasts That Were Rare Pre-Treatment

In post-treatment bone marrow from patient F003, discrete populations of persisting cells made up the majority of the sample (80%). In order to perform a more direct comparison of pre- and post-treatment AML blasts, these blast populations were gated from Day 0 and Day 14 bone marrow and remapped in viSNE. They were then visualized in order to demonstrate subpopulations based on cellular abundance (Figure 3-6A). Thirteen AML cell populations were identified based on phenotype and abundance in pre-treatment and post-treatment samples. Of these AML cell populations from patient F003, 6 populations were abundant pre-treatment and 7 totally distinct populations were abundant post-treatment (Figure 3-6A).

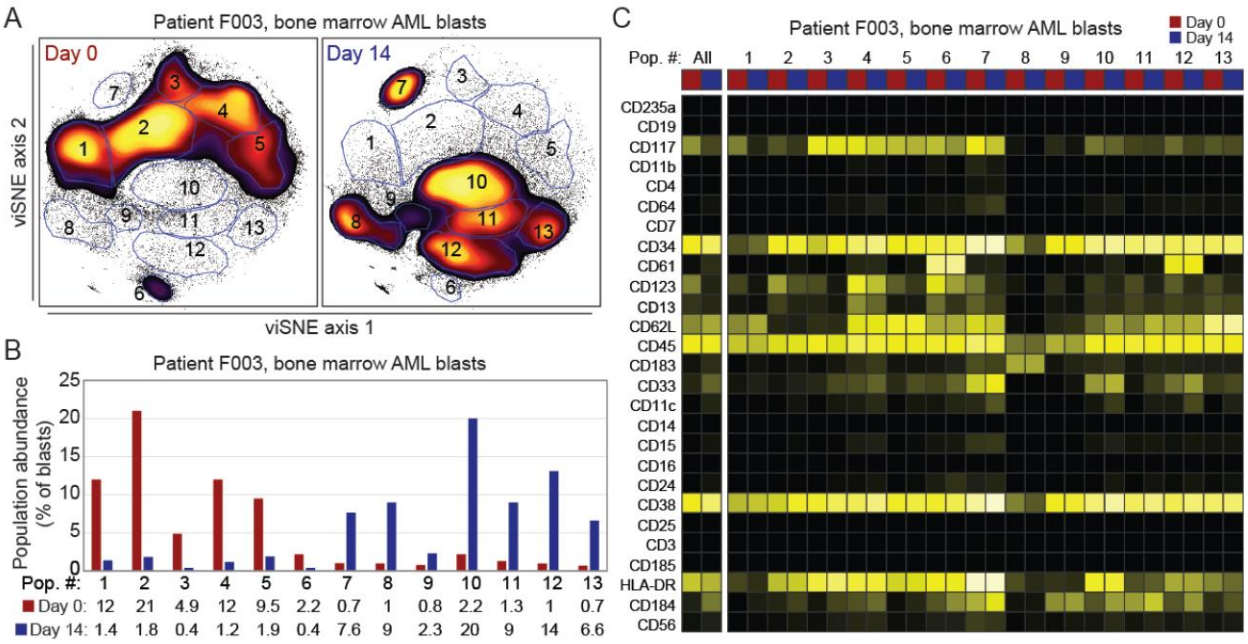

Figure 3-6. Rare subsets at diagnosis become prominent after treatment in a patient with refractory AML. (A) Blasts from Day 0 and Day 14 were gated out from prior viSNE maps for patient F003 (shown in Figure 3-4) and remapped in viSNE together. Gates were drawn around subpopulations and based on relative cell abundance, as in Figure 3-1B. (B) Percent of total blasts for each gate at both time points. (C) A heat map of median marker expression for each gate at both time points is shown.

Overall, the AML phenotype shifted markedly post-treatment. Over 85% of cells post-treatment were groups into 7 phenotypic regions as mentioned above. By contrast, fewer than 10% of the cells fell into these regions pre-treatment (Figure 3-6B). The subpopulations observed post-treatment ranged in abundance from 2.3% to 20% of the AML cells. At diagnosis, in the pre-treatment sample, these populations constituted as little as 0.6% to 1.8% of total blasts (Figure 3-6B).

Increased Expression of CD34, CD38 and CD184 Characterized Post-Treatment Persistent AML While Subpopulations Displayed Significant Heterogeneity

Individual marker changes contributing to overall shift were then characterized. Several markers of leukemia stem cells were identified as contributing to the overall shift in phenotype in samples from patient F003 (Figure 3-6C and Figure 3-9). Median marker expression within the 13 identified subsets was calculated and visualized as a heat map (Figure 3-6C). Prominent changes seen in the pre- to post-treatment samples include both overall increases and decreases in specific markers. Among these markers were CD34, CD38, and CXCR4/CD184, which respectively increased by 0.7, 0.9, and 0.6 fold on the log-like asinh_{15} scale. Expression of CD117 and CD123, markers associated with LSC ability, decreased in the post-treatment sample. In non-leukemic cells,

expression of CD34, CD38, and major cell type identity markers (e.g. HLA-DR, CD4, CD19) did not significantly change over time on non-AML cells (all <0.2 fold, Figures 3-7 and 3-8).

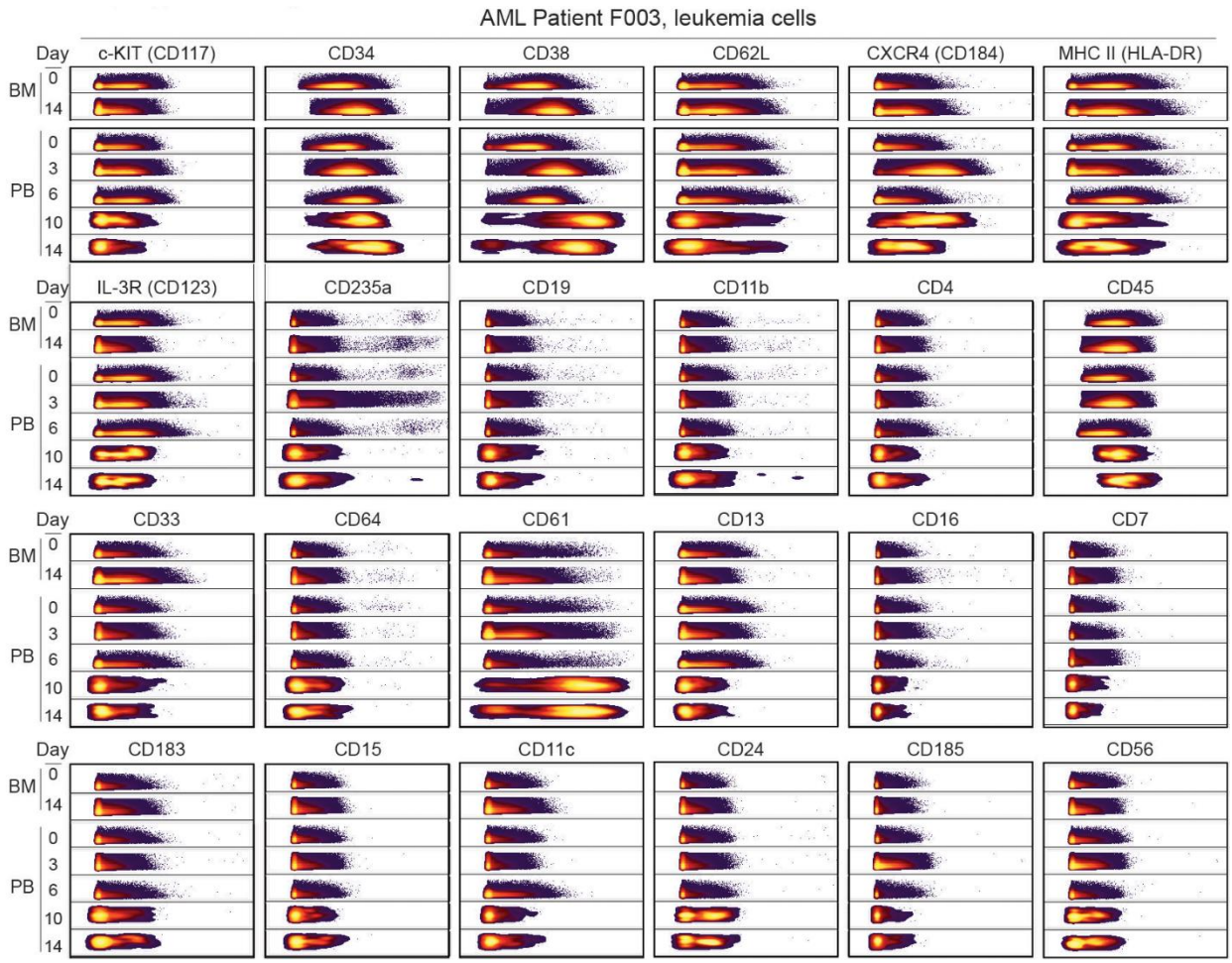

Figure 3-7 - Changes in individual markers over time during treatment on AML blasts from patient F003. Biaxial plots summarize six clinical timepoints (rows) for 24 markers (sets) for the AML blast cells from patient F003, gated as shown in Figure 3-4. The indicated marker is plotted on the x-axis using the same arcsinh_{15} scale as in other figures (e.g. Figure 3-1B). Plot labels are omitted to save space. The y-axis is mass cytometry event length, which is used here to spread the events out in the y-axis to create a compressed band plot view that allows rare subsets to be observed (see e.g. CD235a) that would be obscured in a traditional 1D histogram view.

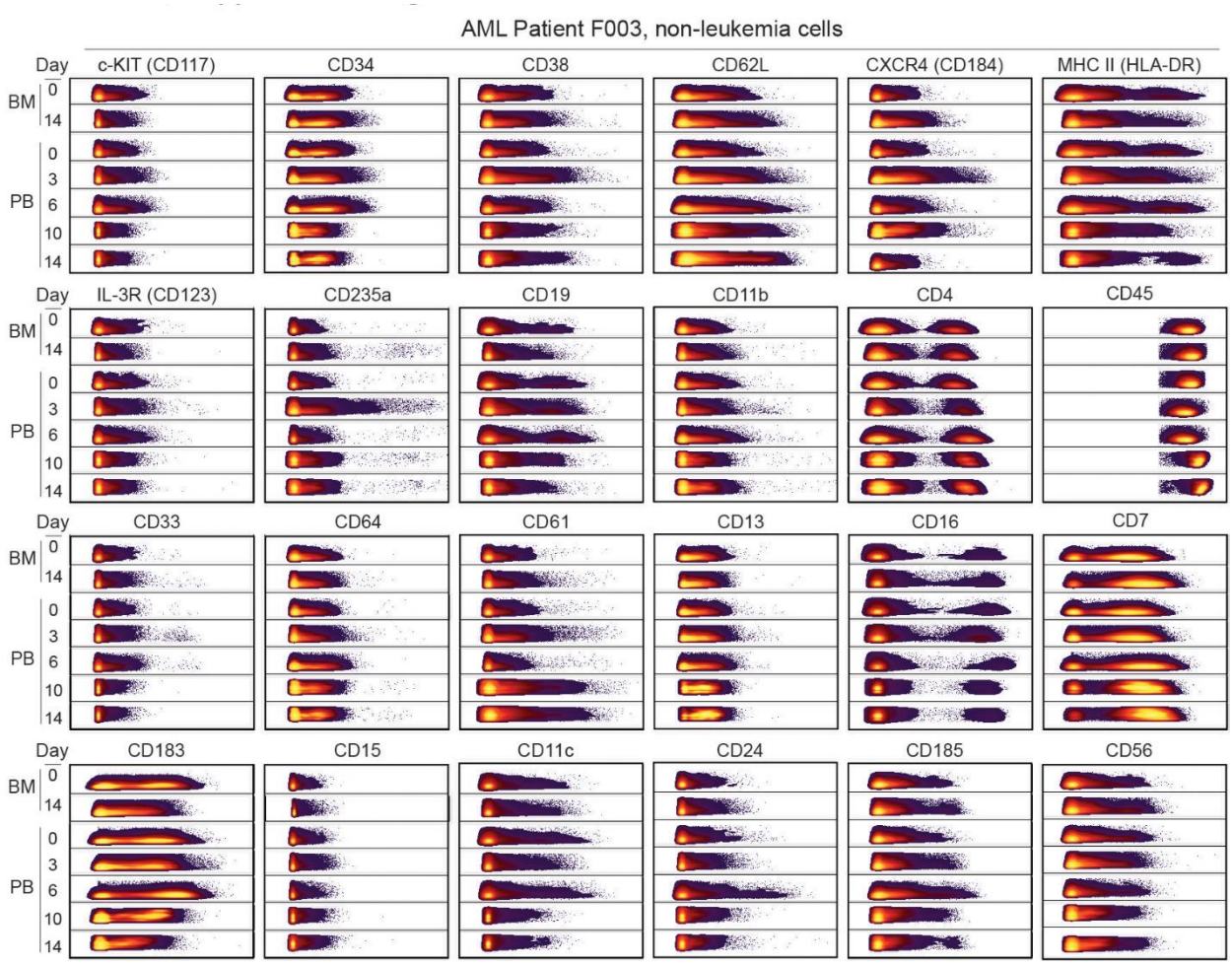

Figure 3-9 - Changes in individual markers over time during treatment on non-leukemia cells from patient F003. As in Figure S2, biaxial plots summarize six clinical timepoints (rows) for 24 markers (sets) for the non-leukemia cells from patient F003, gated as everything not in the leukemia blast gate shown in Figure 3-4. The indicated marker is plotted on the x-axis using the same arcsinh_{15} scale as in other figures (e.g. Figure 3-1B). Plot labels are omitted to save space. The y-axis is mass cytometry event length, which is used here to spread the events out in the y-axis to create a compressed band plot view that allows rare subsets to be observed (see e.g. CD16) that would be obscured in a traditional 1D histogram view.

Analyses limited to overall changes in the entire population of persistent cells belie the underlying heterogeneity seen in individual subpopulations of cells, where changes are quite diverse. Though CD117/*c-kit* expression showed a large decrease from initial sample to post-induction sample, the heat map and biaxial plots demonstrate that expression of this marker is heterogeneous and remains high in one subpopulation (population 7 in Figure 3-6C and Figure 3-9). Subpopulation 7 is also notable for increased expression of CD33 and CD184/CXCR4, which are both quite low in all of the prominent pre-treatment populations (Figure 3-9). Based on this, it is apparent that a heterogeneous group of subpopulations constitute AML before and after treatment.

Thus, the overall shift in AML phenotype was driven by the differential abundance of phenotypically distinct subpopulations that were rare pre-treatment (Figure 3-6A-C).

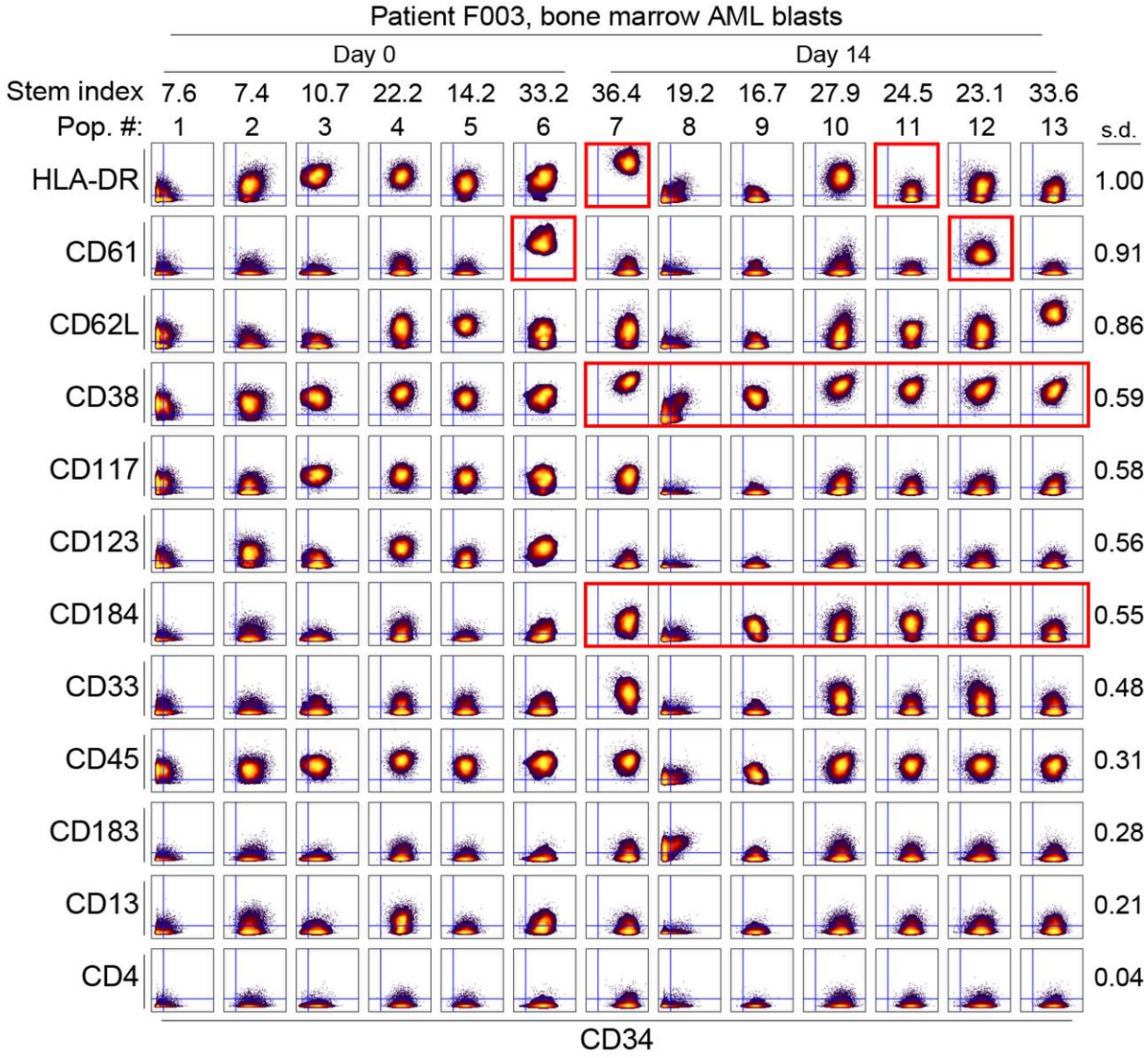

Figure 3-9. Single cell analysis of immunophenotype in AML subpopulations before and after therapy response. Biaxial density plots show the markers with the highest standard deviation across subsets compared with CD34 on individual cells in live bone marrow AML blasts from patient F003. Columns show AML blast populations identified (Figure 3-6) from Day 0 or Day 14, whichever was more abundant. Red boxes highlight key changes in protein expression discussed in the text. Standard deviation (SD) of each row's marker across all populations is indicated to the right of the plots. Stem index (Figure 3-2) of each population is indicated above each column. CD34 and the top 11 most variable markers were graphed. CD4 was included as an example of a low expression marker that did not change.

Marker Expression Varied Among Subpopulations

To quantify variability within subpopulations, we calculated the standard deviation of each marker among all 13 subpopulations identified in pre- and post-treatment samples. The most variable markers among the subpopulations are shown in Figure 3-9, with the exception of CD4, which was a negative, non-variable marker on these leukemia subpopulations. HLA-DR displayed the most variability (SD=1.0), followed closely by CD61 (SD=0.91) and CD62L (SD=0.86). Expression of these markers ranged from high expression to little or no expression across the 13 populations, with variability seen in both pretreatment and post-treatment samples. Interestingly, population 8 was identified as having little or no expression of most markers that identified leukemic blasts, though this population was seen in both pre-and post-treatment samples and its constituent cells were identified in the initial gating schemes (Ir+, single cells).

Persisting AML Cells Become Less Phenotypically Stem-Like

In order to understand how individual marker changes affected overall phenotypic diversity and identity with respect to known, healthy populations, pretreatment (Day 0) and post-treatment populations (Day 14) blast cells were mapped in viSNE with the normal healthy marrow population. Based upon the resultant viSNE map, phenotypic shifts were quantified and AML blasts were shown to be less phenotypically similar to healthy HSCs following treatment (Figure 3-10B). For this analysis, predominant bone marrow blast populations at diagnosis (AML subsets 1 - 6) were mapped with normal marrow and major persisting populations from day 14 (subsets 7 - 13) (Figure 3-10A&B). Post-treatment, the mean change in HSC distance was +10.0, indicating a phenotypic shift away from HSCs (Figure 3-10B). Taken together, these data indicate that the overall trend is for AML blasts to gain a novel phenotype that is distinct from those of healthy stem and progenitor cells in the bone marrow.

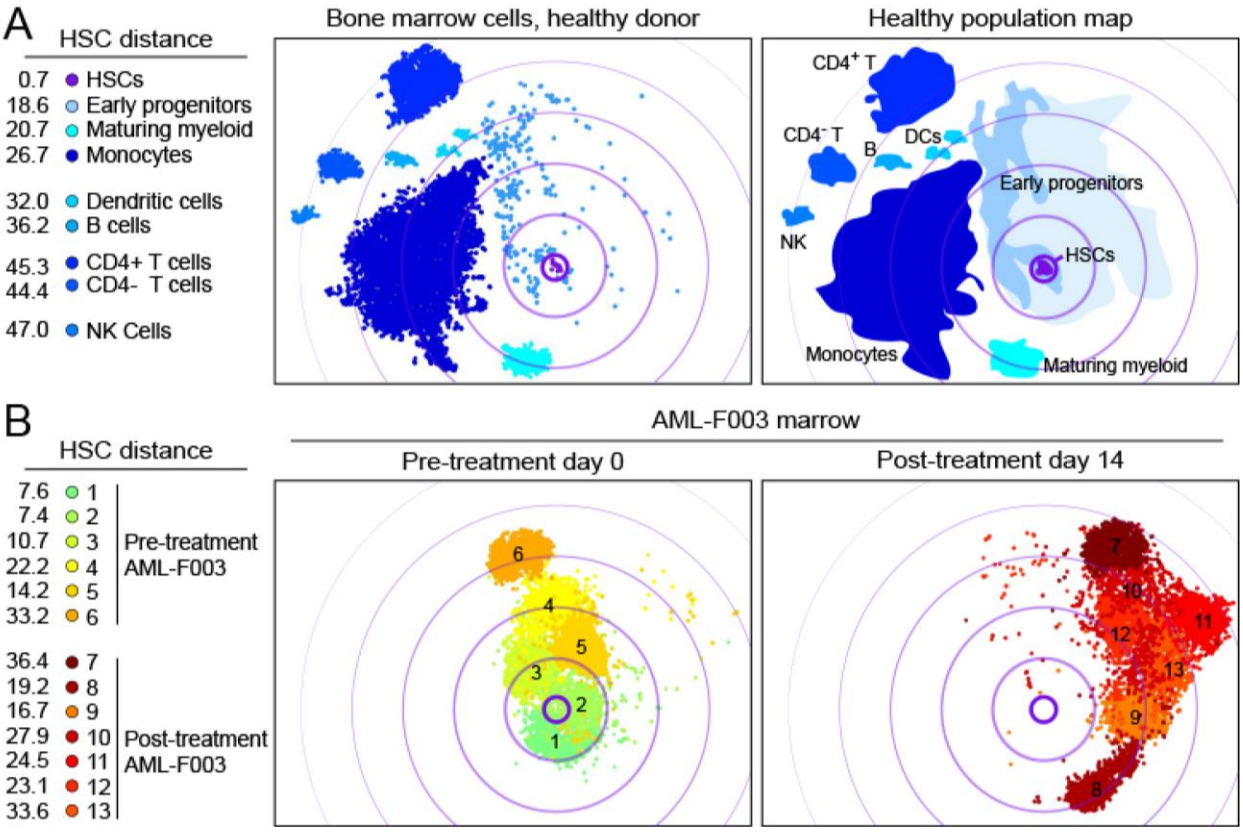

Figure 3-10. Blast populations become increasingly phenotypically different from both differentiated cells and non-malignant hematopoietic stem cells. 27-dimensional viSNE analysis compared normal bone marrow cells, AML blasts from Day 0 from patient F003 (subpopulations 1-6 only), and AML blasts from Day 14 from patient F003 (subpopulation 7-13 only). (A) Normal bone marrow mononuclear cells are shown (left). A cartoon outline of the major healthy subpopulations was identified and HSC distance measured (right). (B) Pre-treatment (day 0) and post-treatment (day 14) samples are compared same map as in (A) and HSC distance of each subpopulation is shown.

Discussion

This work characterizes changes in AML blasts and healthy cells over time following induction chemotherapy according to 27 protein features, including the vast majority of diagnostic markers for this disease. A key benefit of this approach is that it reveals whether AML cells are cleared, phenotypically altered, or left unchanged after treatment. This approach directly measures the impact of treatment on the cellular milieu of the patient's marrow and blood. By identifying persistent AML cells post therapy, mass cytometry and computational analysis provided a clear picture of AML subpopulation dynamics in the early therapy response and revealed the unexpected finding that AML "persister" cells can become significantly less phenotypically stem-like immediately following treatment. These results provide a foundational reference for understanding the reservoir of AML cells

that evade treatment and provide a proof of concept for high dimensional single cell characterization of clinically relevant cell-surface molecules in AML.

While this study assessed samples from 5 patients, there are 46 samples in total that have been analyzed. Regardless, this pilot study assessed the feasibility of greatly expanded longitudinal monitoring of AML to capture differentiation and cell subset dynamics. AML displays significant phenotypic plasticity in the clinical setting (Baer, Stewart et al. 2001). However, this phenotypic heterogeneity is not well represented in focused fluorescent cytometry panels measuring between six and eight features. Here we do not intend to duplicate work done by large clinical flow consortia, such as EuroFlow, rather we intend to demonstrate the utility and feasibility of this approach (van Dongen, Lhermitte et al. 2012, van Dongen, Orfao et al. 2012). Larger studies of high-dimensional immune cell monitoring in AML are now needed to further validate this approach and to begin to understand treatment specific changes. This study demonstrates potential drawbacks of small fluorescence cytometry panels used to distinguish the leukemia blasts at during therapy – namely that changes in certain markers may be missed by small panels. The results of this study also indicate that immunophenotype in AML is highly plastic and can fundamentally shift in a matter of days in response to treatment to adopt a pattern that is both abnormal and not HSC-like. Furthermore, while this study was not designed to determine whether or not chemotherapy is altering cell phenotypic expression or selecting for persistent cells, the kinetics of early marker change in the peripheral blood indicate that phenotypic plasticity is possible.

This focused pilot study indicates that mass cytometry is feasible for clinical immune monitoring of treatment impact in blood cancers, such as AML. Furthermore, these findings are immediately translatable to longitudinal studies of other hematologic malignancies, chronic viral infections, and autoimmune disorders. Early phenotypic changes may also be evaluated in the peripheral blood blasts, when present, and could give key insight to early treatment success. It has been shown that early blasts decline in peripheral blood can predict complete remission rates in AML (Gianfaldoni, Mannelli et al. 2006, Elliott, Litzow et al. 2007, Arellano, Pakkala et al. 2012). With added information of high-dimensional directionality and phenotypic distance changes, peripheral blood changes can be better characterized early and could become even more informative. However, because peripheral blasts can have a slightly different immunophenotype than bone marrow blasts (as in F001), changes in blasts post-therapy are likely best evaluated within the same compartment. Visualization of high-dimensional relationships

within viSNE allow for assessment of changes over time as well as noting any differences between different compartments (e.g., marrow vs. blood). The approach presented here also provides a new clinical tool for tracking and dissecting cellular identities and functional responses in clinical studies and translational research. For example, the markers identified here can now be used to sort persister-like cells from pre-treatment AML and determine whether these cells are genetically related to the persister cells. Recently published data reveals the decoupling of immunophenotype and signaling in AML samples at diagnosis and shows that the presence of "primitive" signaling profiles found in healthy HSCs likely more prognostic than phenotype (Levine, Simonds et al. 2015). Applying this finding to our data, we would predict that refractory AML adopts a more "primitive" signaling pattern post-treatment, regardless of phenotypic changes. In this way, application of signaling measurements to longitudinal monitoring in AML has great potential to reveal signaling and phenotypic markers of resistance that could improve disease targeting.

This study describes a novel phenotypic stem-ness index based on the high dimensional immunophenotype of normal and leukemia samples. It is possible that the majority of post-treatment cells are leukemia stem cells and their phenotypic changes reveal a level of plasticity in stem cell phenotype that cannot to be assessed outside of real world treatment situations. Importantly, some of this plasticity has been seen in vitro as well (Eppert, Takenaka et al. 2011, Sarry, Murphy et al. 2011). It is possible that regardless of their "stem-ness", as defined either by phenotype or function, these therapy resistant cells are the cells that warrant further investigation and therapeutic focus in the context of AML, as it is their persistence post-treatment that results in the ultimate death of patients. We have shown that subsets present in a refractory AML display increased expression of CD34, CD38, and CD184 while viSNE analysis shows that these cells "move away" from healthy stem cells and the bulk of the original disease. These cells went in a new direction immunophenotypically, maintaining many AML cell characteristics while also adopting new features. Furthermore, this could indicate some subtle maturation, however abnormal, within the blast population induced by the treatment. It would not be completely unexpected for LSCs to switch immunophenotype, as though LSCs were first described as Lin^{-} $CD34^{+}$ $CD38^{lo/}$, it is well-established that cells with leukemia initiating ability and stem-ness properties exist outside of this original definition (Taussig, Miraki-Moud et al. 2008, Taussig, Vargaftig et al. 2010, Goardon, Marchi et al. 2011, Sarry, Murphy et al. 2011). This may be especially true in cases of refractory AML.

Additionally, the expression of CD184 has been linked to worse outcomes in AML and therapeutic targeting of this molecule has been reported (Spoo, Lubbert et al. 2007, Nervi, Ramirez et al. 2009).

This refractory AML presented here demonstrated a large burden of persistent leukemia cells after therapy. These persister cells are common in AML, as evidence by MRD in patients who achieve a remission or as overt refractory disease. At no point did we see AML "collapse" into a phenotypically more similar (less heterogeneous) population or adopt a uniform stem or progenitor cell phenotype that could imply a bottleneck selection event. The response to treatment was typically an increase in diversity of phenotypes and movement away from traditional stem cell phenotypes. It may be that there is an inducible mechanism of phenotypic shift in AML cells caused by chemotherapy. This mechanism could depend on non-cell autonomous changes and inter-clonal interaction, as has recently been described in cancer cells (Marusyk, Tabassum et al. 2014). This would contrast with the classic "selecting for a subclone" model and aligns with some recent observations based on gene sequencing, which highlight genetic and functional heterogeneity within AML at diagnosis and relapse (Ding, Ley et al. 2012, Klco, Spencer et al. 2014). Further clarification of the functional differences in immunophenotypically distinct subsets of AML, particularly of intracellular signaling networks, is needed to identify novel therapeutic targets in this disease.

Acknowledgements

This study was supported by NIH/NCI K12 CA090625 (P.B.F), R25 CA136440-04 (K.E.D.), R00 CA143231-03 (J.M.I.), the Vanderbilt-Ingram Cancer Center (VICC, P30 CA68485), and a VICC Hematology Helping Hands award.

Contributions:

Planning (JMI, PBF, SRM), sample acquisition and processing (PBF, HGP, JMI, ACS), mass cytometry panel design (PBF, JMI, ACS), experiment design (PBF, JMI), mass cytometry experimentation (PBF), data analysis (PBF, KED, JMI), and figures and writing (PBF, JMI).

CHAPTER 4

CHARACTERIZING CELL SUBSETS IN HETEROGENEOUS TISSUES USING MARKER ENRICHMENT MODELING

Authors: K. E. Diggins, A. R. Greenplate, N. Leelatian, C. E. Wogsland, and J. M. Irish

Preface

Methods such as those introduced in Chapter 1 and Chapter 2 are useful for discovering phenotypically distinct cell subpopulations in high-dimensional, single-cell data. However, I found the characterization of computationally discovered subsets to be largely qualitative and subjective. There was no automated method for quantitatively comparing the populations that had been discovered using biaxial gating or different computational tools. To address this need, I developed an equation that combines magnitude and variance information to determine context-specific feature enrichment on a cell subset compared to a reference cell subset. This equation is the core of the method termed Marker Enrichment Modeling (MEM) that takes previously identified cell population data as input and reports a label for each subset relative to a specified reference cell population. Using multiple flow cytometry datasets from the analysis of human and murine tissues, I illustrated the ability of MEM to quantify proteins on cell subsets discovered across analysis methods, platforms, and tissues. The MEM equation and concept is a key component to the long-term goal of developing fully automated analysis workflows that can identify populations of cells from high dimensional data and subsequently characterize and label those cells without the need for expert interpretation.

Abstract

While advances in single cell biology have enabled measurement of 35 or more features on millions of single cells, learning cell identity still relies on human experts. *Marker Enrichment Modeling* (MEM) addresses the need to objectively describe cell subsets by quantifying feature enrichment and reporting a text label that can be used by machines and humans to learn cell type (cytotype). In datasets including healthy human blood, bone marrow, and tonsil, murine tissues, and human tumors, MEM identified key proteins used by experts to

distinguish rare and novel cell subsets. MEM-generated labels accurately characterized phospho-protein signaling responses in subgroups of leukemia patients. Algorithmic comparison of MEM label correctly identified 80 cell populations from 7 studies of 3 human tissues measured using different instrumentation and distinguished tumor-infiltrating immune cell subsets and malignant cell populations in human glioma. MEM labels thus provide a quantitative language to objectively communicate characteristics of new and established cell types observed in complex tissue microenvironments.

Introduction

Quantitative cytometry workflows have developed diverse approaches to grouping cells into populations(Lo K 2009, Qiu, Simonds et al. 2011, Shekhar, Brodin et al. 2013, Bruggner, Bodenmiller et al. 2014, Mosmann, Naim et al. 2014) and visualizing results in graphs that arrange populations based on phenotype(Irish, Hovland et al. 2004, Bendall, Simonds et al. 2011, Bendall, Davis et al. 2014, Levine, Simonds et al. 2015, Spitzer, Gherardini et al. 2015). Significant features of populations are typically assumed to be those which were expressed at the highest median level on each population. These strategies work well when feature variability is low and cells match established types, but computational analysis of single cell data is now routinely revealing novel cells with non-canonical phenotypes(Becher, Schlitzer et al. 2014, Irish 2014, Patel, Tirosh et al. 2014, Greenplate, Johnson et al. 2016). This is especially the case for diseases where abnormal expression profiles and signaling responses distinguish clinically significant cell subsets(Irish, Hovland et al. 2004, Irish, Myklebust et al. 2010, Levine, Simonds et al. 2015, Ferrell, Diggins et al. 2016, Greenplate, Johnson et al. 2016). Existing statistical approaches, such as a z-score, can be used to characterize a population's deviation from a mean, but may be limited to a normal distribution or may not account for both intra- and inter-population variability in a single metric.

We have developed Marker Enrichment Modeling (MEM) to quantify context-specific marker enrichment on populations. MEM was developed to address the idea that the most experimentally and biologically relevant cellular features are those that are more homogenously expressed on the majority of cells in a given population than on other cells in the sample or experimental framework. Additionally, features that are specifically lacking on a population (i.e. negatively enriched) are as important to that population's identity as are those specifically

expressed at high levels (i.e. positively enriched). Given these considerations, the MEM equation was derived to produce a signed value for each population feature that quantifies positive and negative, population-specific, contextual feature enrichment. For a given subset of cells within a sample, MEM compares both feature magnitude and feature variability for the subset with those of a reference cell population, such as stem cells or other cells in the samples. MEM quantifies the extent to which proteins or other measured features are enriched on a chosen cell subset compared to a selected reference population. For each measured feature, a MEM score is calculated (Eq. 1).

$$\text{MEM score} = \left| MAG_{POP} - MAG_{REF} \right| + \left(\frac{IQR_{REF}}{IQR_{POP}} \right) - 1, \quad (MAG_{POP}\text{-}MAG_{REF}) < 0 \rightarrow \text{MEM} = \text{-MEM} \qquad \text{(Eq. 1)}$$

In Eq. 1, POP denotes the population of interest, REF denotes the reference population to which POP will be compared, MAG is the magnitude of a feature (here, median protein expression detected by mass or fluorescence flow cytometry), and IQR indicates the interquartile range. The ratio in IQR values is used to capture how homogeneously the chosen cells in POP expressed a given protein relative to REF. The reference population REF is chosen based on the biological comparison of interest (Figure 4-1). Examples of REF include all other cells from a sample not in POP Figure 4-2, and Figure 4-13), previously-defined populations, such as other cells from the sample (Figure 4-6 and Figure 4-9), comparisons to cells from other individuals (Figure 4-15), and “gold standard” populations like stem cells (Figure 4-13, Figure 4-14). Unsupervised reference selection (i.e. using all non-population cells from the same sample or set of samples as POP) is advantageous when labeling and characterizing cell systems or microenvironments in the original biological context. Comparison to a common reference facilitates comparing changes in phenotype or sample quality over time, quantifying differences between individuals, and matching a label from a new population of cells to established reference populations of diseased or healthy cells.

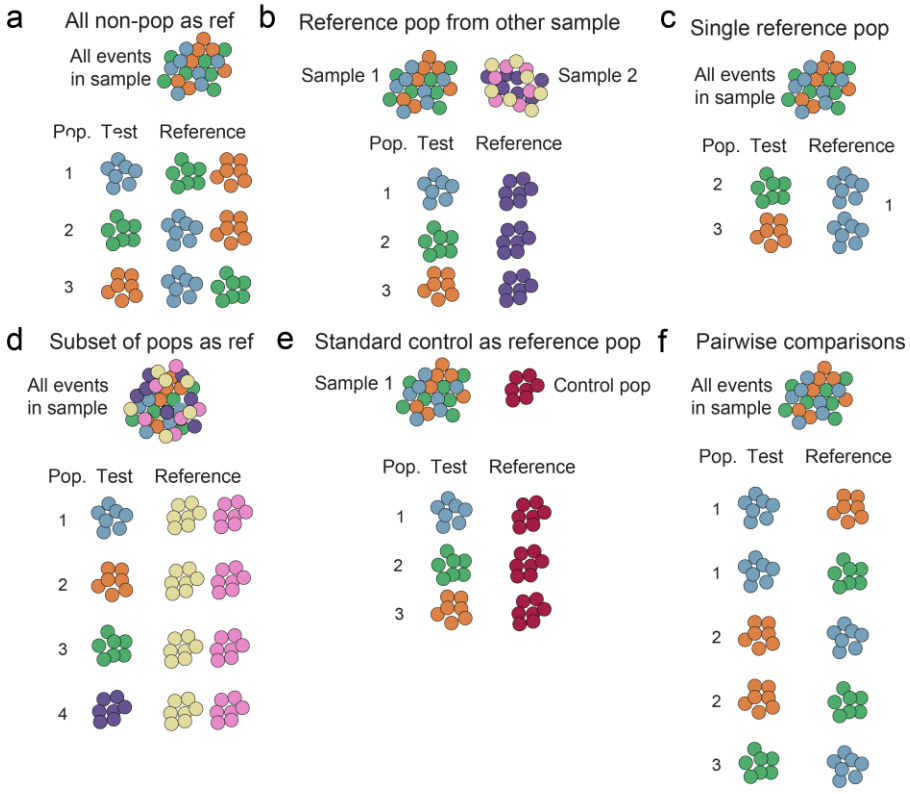

Figure 4-1. Examples of MEM reference population selection to capture different contexts. Alternative reference populations (REF) can be used to capture how features of the test population (POP) are enriched in different contexts. Reference comparisons include a) all non-population cells in the sample or experiment (default), b) a population from another sample in the same study, c) a population from the same sample, d) multiple subsets of non-population cells from the same sample, e) a standard control population, and f) pairwise comparison between all populations in a sample.

MEM arranges the scores calculated for each protein into a quantitative label for each POP. For example, the MEM label for $CD4^+$ T cells from healthy human tonsils identified by expert gating of mass cytometry data(Polikowsky, Wogsland et al. 2015) and compared to all other immune cells in the sample as REF was ▲$CD4^{+10}$ $CD3^{+9}$ $CD5^{+5}$ $CD27^{+5}$ ▼$MHCII^{-10}$ $CD19^{-9}$ $CD20^{-7}$ $CD40^{-7}$ $CD22^{-7}$ $Ig\kappa^{-5}$ $Ig\lambda^{-4}$ IgD^{-2} $SHP1^{-2}$ $CD16^{-2}$ $CD33^{-2}$ IgM^{-2} (Table 4-1). In this MEM label, the exponent quantifies protein enrichment in tonsillar $CD4^+$ T cells relative to the other the tonsillar leukocytes, primarily B cells. This MEM label indicates that tonsillar $CD4^+$ T cells are distinguished both by expression of canonical T cell proteins ($CD3^{+9}$ $CD5^{+5}$) and by a lack of expression (negative enrichment) of antigen presenting cell (APC) and B cell proteins ($MHCII^{-10}$ $CD19^{-9}$ $CD20^{-7}$). MEM scores are robust across tissue types and experiments; $CD4^+$ T cells were consistent in MEM labels for the same tissue across healthy donors and correctly captured differences in phenotype between tissues (Table 4-1). For example, tonsillar $CD4^+$ T cells displayed significantly greater CD27 enrichment than blood $CD4^+$ T cells ($p<0.02$; tonsillar $CD4^+$ T cells: CD27▲5.3 ± 0.58, N=3; blood $CD4^+$ T cells: CD27▲3.4 ± 0.98, N=7; see Table 4-1).

Highly enriched proteins were more important to accurate population identification than proteins with high median expression (Figure 4-2c; Figure 4-4; Figure 4-5).

MEM label reporting provides a new way to objectively and quantitatively communicate cell type in studies of complex, dynamic tissue microenvironments, such as stem cell niches, immune responses, and human tumors.

Table 4-1 – Healthy human CD4$^+$ T cells from various mass cytometry studies were labeled consistently by MEM

Tissue	Study	Panel focus	#	MEM label for CD4$^+$ T cells vs. other live cells
Bone marrow	Bendall et al.(Bendall, Simonds et al. 2011)	Hematopoiesis	4	▲CD4^{+10} CD3^{+6} ▼MHCII^{-3} CD8^{-2} CD45RA^{-2} CD11b^{-2}
	Amir et al.(Amir el, Davis et al. 2013)	Canonical immune	5	▲CD4^{+10} CD3^{+4} ▼CD45RA^{-2}
	Ferrell et al. (Ferrell, Diggins et al. 2016)	AML & myeloid cells	2	▲CD3^{+10} CD7^{+7} CD4^{+6} CD62L^{+4} ▼CD11b^{-6} CD11c^{-6} MHCII^{-6} CD64^{-5} CD61^{-4} CD13^{-3} CD38^{-3} CD123^{-2} CD33^{-2} CD14^{-2}
Healthy human PBMCs	Leelatian et al.(Leelatian, Diggins et al. 2015)	Canonical immune	7	▲CD4^{+10} CD3^{+7} ▼CD8a^{-3} CD16^{-3} CD11b^{-2} CD69^{-2} MHCII^{-2}
	Greenplate AR[§]	T cells	1.3	▲CD4^{+10} CD3^{+9} CCR7^{+5} CD27^{+5} CD28^{+3} CD64^{+2}
			1.7	▲CD4^{+10} CD3^{+9} CD45RO^{+3} CD27^{+3} CD64^{+2} CD28^{+2} ▼CD45RA^{-4}
			1.1	▲CD4^{+10} CD3^{+8} CD27^{+4} CCR7^{+3} CD64^{+2} ▼CD45RA^{-2}
			1.8	▲CD4^{+10} CD3^{+9} CD45RO^{+4} CD27^{+3} CD43^{+2} ▼CD45RA^{-5}
			1.6	▲CD4^{+10} CD3^{+7} CD27^{+2} ▼CD45RA^{-4}
			1.4	▲CD3^{+10} CD4^{+9} CD27^{+4} CCR7^{+3} CD28^{+2} ▼CD45RA^{-3}
			1.2	▲CD4^{+10} CD3^{+10} CD43^{+4} CD27^{+3} CCR7^{+2} CD28^{+2} CD45RO^{+2} ▼MHCII^{-4} CD38^{-2} CD45RA^{-2}
Healthy human tonsil	Polikowsky et al.(Polikowsky, Wogsland et al. 2015)	B cells	3.1	▲CD4^{+10} CD3^{+9} CD5^{+5} CD27^{+5} ▼MHCII^{-10} CD19^{-9} CD20^{-7} CD40^{-7} CD22^{-7} Igκ^{-5} Igλ^{-4} IgD^{-2} SHP1^{-2} CD16^{-2} CD33^{-2} IgM^{-2}
			3.4	▲CD4^{+9} CD3^{+9} CD5^{+5} CD27^{+5} ▼MHCII^{-10} CD19^{-10} CD20^{-9} CD22^{-7} CD40^{-7} Igλ^{-3} Igκ^{-3} IgG^{-2} CD33^{-2} CD16^{-2} SHP1^{-2}
			3.5	▲CD4^{+10} CD3^{+10} CD27^{+6} CD5^{+4} ▼MHCII^{-7} Igλ^{-5} CD20^{-5} CD19^{-5} CD22^{-5} CD40^{-4} IgD^{-3} IgM^{-3} CD16^{-2} Igκ^{-2} CD79B^{-2} SHP1^{-2}

Sample numbers match study numbers in Figure 4-11.

[§]Unpublished data from AR Greenplate, prepared as in Leelatian et al.(Leelatian, Diggins et al. 2015)

Results

Four cytometry studies, Dataset A from Leelatian et al.(Leelatian, Diggins et al. 2015), Dataset B from Bendall et al.(Bendall, Simonds et al. 2011), Dataset C from Becher et al.(Becher, Schlitzer et al. 2014), and Dataset D containing unpublished human glioma data(Leelatian, Doxie et al. 2016) were used to evaluate MEM and identify biological features of expert and machine identified cell subsets. These datasets had been previously analyzed by experts and by computational tools including viSNE(Amir el, Davis et al. 2013) and SPADE(Qiu, Simonds et al. 2011), which are respectively used in mass cytometry for dimensionality reduction and cell clustering(Diggins, Ferrell et al. 2015).

Dataset A was mass cytometry data quantifying expression of 25 proteins on healthy human peripheral blood mononuclear cells (PBMC)(Leelatian, Diggins et al. 2015). This dataset was chosen for two reasons: 1) the 7 cell subsets present are well-established and phenotypically distinct populations that served as a gold standard of biological 'truth' and 2) the cells in each of the 7 subsets had been characterized for 25 proteins that displayed varying homogeneous and heterogeneous expression patterns across the subsets. Populations were expert gated following viSNE reduction to a two-dimensional view and each population was compared to the other cells in the sample as reference (Figure 4-2). MEM returned automatic labels that matched prior expert analysis(Leelatian, Diggins et al. 2015) and correctly assigned high positive enrichment values to canonical protein features of each subset (Figure 4-2b), including CD4 on $CD4^+$ T cells (▲$CD4^{+6}$ $CD3^{+5}$ ▼$CD8a^{-4}$ $CD16^{-3}$), IgM on IgM^+ B cells (▲MHC II^{+8} IgM^{+6} $CD19^{+5}$ ▼$CD4^{-6}$ $CD3^{-5}$), CD11c and MHC II on monocytes (▲$CD11c^{+8}$ $CD33^{+7}$ $CD14^{+6}$ $CD61^{+6}$ MHC II^{+4} $CD44^{+3}$ ▼$CD3^{-5}$ $CD4^{-4}$), and CD16 on NK cells (▲$CD16^{+9}$ $CD56^{+2}$ $CD11c^{+2}$ ▼$CD4^{-7}$ $CD3^{-4}$ $CD44^{-3}$). Proteins that were not significantly enriched in any of the 7 subsets of mature human blood mononuclear cells were correctly assigned near-zero MEM scores (e.g. CD34 and CD117 proteins normally expressed on hematopoietic stem cells, Figure 4-2b). Similarly, proteins with little variability across cell subsets were assigned low, near-zero MEM scores, even for the proteins with the highest median expression values (e.g. CD45 on all subsets or CD45RA on non-T cell subsets Figure 4-2b).

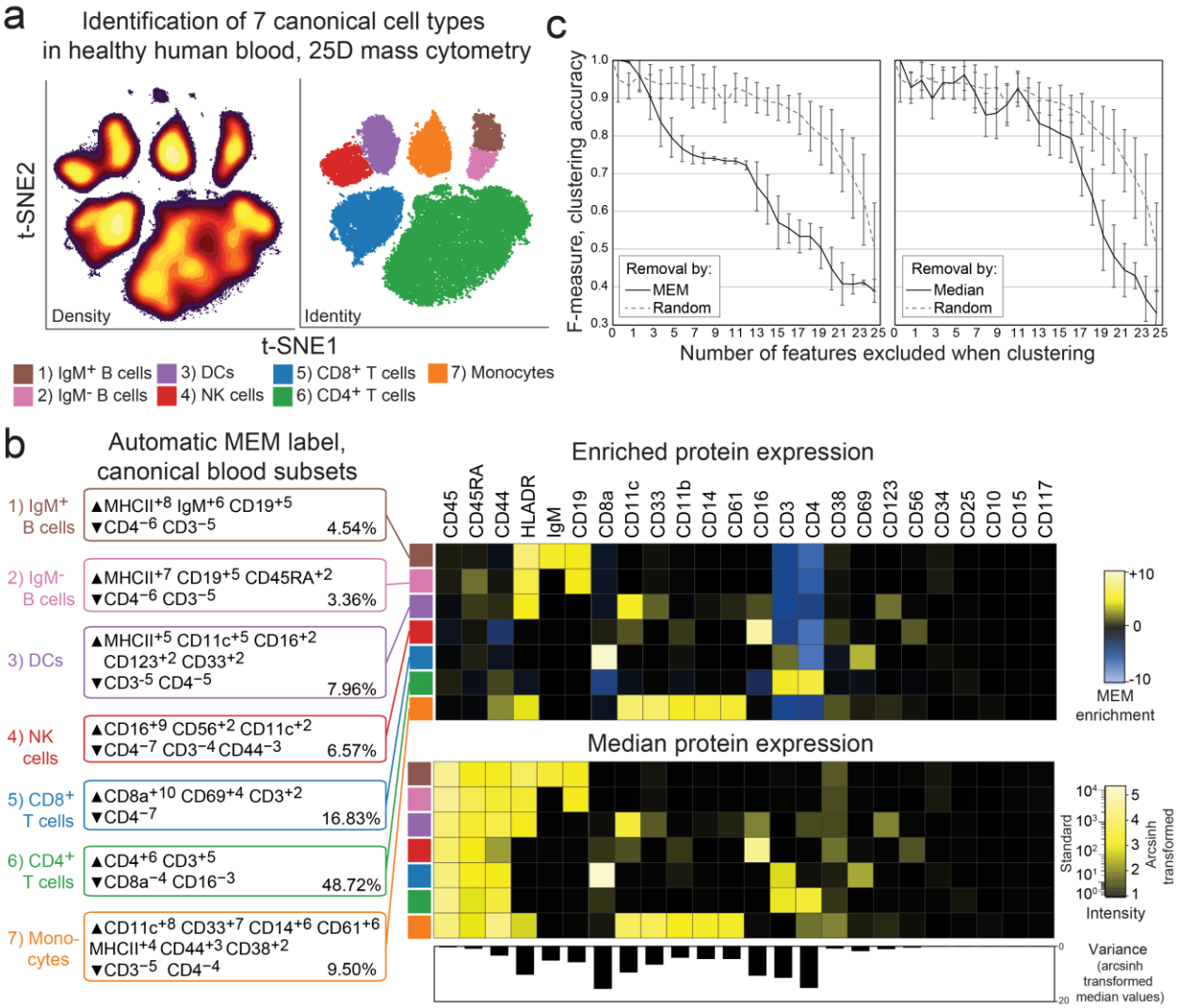

Figure 4-2: Marker enrichment modeling (MEM) automatically labels human blood cell populations in Dataset A. a) Cells from normal human blood were previously grouped into 7 canonical populations using viSNE analysis and expert review of 25D mass cytometry data(Leelatian, Diggins et al. 2015). b) MEM labels were computationally generated for each canonical cell subset using the other six populations as reference. The population labeled by immunologists as "CD4^{+} T cells" was labeled by MEM as ▲CD4^{+6} CD3^{+5} ▼CD8a^{-4} CD16^{-3} and comprised 48.72% of PBMC in this sample. In contrast, the MEM label ▲CD16^{+9} CD56^{+2} CD11c^{+2}▼CD4^{-7} CD3^{-4} CD44^{-3} was generated for the population gated as "NK cells". Heatmaps show protein enrichment values used to calculate MEM labels and median protein expression values for each protein on each cell subset. Variability in protein expression across the 7 canonical cell populations is shown below to highlight proteins that were expressed homogeneously (low variability, e.g. CD45) and those that were expressed heterogeneously (high variability, e.g. CD8a, CD4). c) Graphs show decreasing f-measure (clustering accuracy) as markers were excluded from k-means cluster analysis based on high to low absolute MEM or median values, compared to random exclusion.

Proteins enriched in a cell subset but rarely expressed on cells in the reference population were assigned positive MEM scores (e.g. CD19 on B cells, CD16 on NK cells, Figure 4-2b). Negative enrichment scores occurred when proteins were commonly expressed in the reference population but consistently lacking on a specific cell subset. For example, CD3 was expressed in a majority of the cells (CD4^{+} and CD8^{+} T cells

comprised 48.72% and 16.83% of the sample, respectively), thus B cells, NK cells, and monocytes were all assigned negative MEM scores for CD3 (Figure 4-2b). Similarly, CD44 was assigned a negative MEM score for NK cells because all other cells in human blood expressed more CD44 per-cell (Figure 4-2b). Thus, negative MEM scores highlighted a specific lack of expression, or negative enrichment, of a protein on a cell subset relative to the rest of the cells within the sample.

In flow cytometry, differences in protein expression between two cell populations are routinely approximated by comparing median signal intensity. This standard comparison of median signal intensity is equal to the "magnitude difference" portion of the MEM calculation ($MAG_{DIFF} = MAG_{POP} – MAG_{REF}$). To compare MEM to this standard, MAG_{DIFF} and MEM values were calculated for each of the 7 major immune cell populations shown in Figure 4-2 and their corresponding reference populations (Table 4-2).

Table 4-2. MEM equation components for PBMC subsets in Fig. 4-2

		CD45	CD45RA	CD44	HLADR	IgM	CD19	CD8	CD11c	CD33	CD11b	CD14	CD61	CD16	CD3	CD4	CD38	CD69	CD123	CD56	CD34	CD25	CD10	CD15	CD117
CD4 T cells	MAG_{POP}	4.3	2.3	3.2	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	2.4	3.1	0.5	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0
	MAG_{REF}	4.0	2.6	2.6	0.1	0.0	0.0	0.8	0.1	0.0	0.4	0.0	0.0	0.2	0.1	0.0	0.8	0.3	0.0	0.1	0.0	0.0	0.0	0.0	0.0
	IQR_{POP}	0.7	1.7	0.9	0.5	0.5	0.5	0.8	0.5	0.5	0.5	0.5	0.5	0.5	0.9	0.6	0.8	0.5	0.5	0.5	0.5	0.6	0.5	0.5	0.5
	IQR_{REF}	0.8	1.1	1.6	2.3	0.5	0.5	4.8	2.2	0.7	1.2	0.5	0.8	3.7	1.8	0.6	1.2	1.4	0.5	0.7	0.5	0.5	0.5	0.5	0.5
	MEM	0.6	0.1	1.9	-5.0	0.0	0.0	-7.4	-4.7	0.7	-2.5	0.0	0.7	-8.8	4.2	4.1	-1.0	-2.7	0.0	-0.8	0.0	0.0	0.0	0.0	0.0
	MAG_{DIFF-S}	0.5	-0.6	1.3	-0.2	0	0	-0.5	-0.2	0	-0.7	0	0	-0.4	4.8	6.4	-0.4	-0.6	0	-0.2	0	0.2	0	0	0
	MAG_{DIFF}	0.3	-0.3	0.6	-0.1	0.0	0.0	-0.3	-0.1	0.0	-0.4	0.0	0.0	-0.2	2.3	3.0	-0.2	-0.3	0.0	-0.1	0.0	0.1	0.0	0.0	0.0
	IQR_{DIFF}	-0.1	0.7	-0.7	-1.8	0.0	0.0	-4.1	-1.7	-0.2	-0.7	0.0	-0.3	-3.2	-0.8	0.0	-0.4	-0.9	0.0	-0.2	0.0	0.1	0.0	0.0	0.0
CD8 T cells	MAG_{POP}	4.1	2.7	2.6	0.0	0.0	0.0	5.1	0.0	0.0	0.2	0.0	0.0	0.0	2.2	0.1	0.2	1.7	0.0	0.0	0.1	0.0	0.0	0.0	0.0
	MAG_{REF}	4.2	2.4	3.1	0.0	0.0	0.0	0.4	0.0	0.0	0.1	0.0	0.0	0.0	1.7	2.6	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	IQR_{POP}	0.7	1.0	0.8	0.5	0.5	0.5	0.5	0.5	0.5	0.6	0.5	0.5	0.5	0.8	0.5	0.6	0.8	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	IQR_{REF}	0.7	1.4	1.3	0.9	0.5	0.5	0.9	0.7	0.5	0.6	0.5	0.5	0.6	2.5	3.0	1.0	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	MEM	-0.2	0.9	-1.3	1.0	0.0	0.0	7.3	0.5	0.0	0.3	0.0	0.0	0.3	3.5	-10.0	-1.7	1.7	0.0	0.0	0.1	0.0	0.0	0.0	0.0
	MAG_{DIFF-S}	-0.2	0.5	-0.9	0	0	0	10	0	0	0.2	0	0	0	1.1	-5.3	-1	3.5	0	0	0.1	0	0	0	0
	MAG_{DIFF}	-0.1	0.3	-0.5	0.0	0.0	0.0	4.7	0.0	0.0	0.1	0.0	0.0	0.0	0.5	-2.5	-0.5	1.7	0.0	0.0	0.1	0.0	0.0	0.0	0.0
	IQR_{DIFF}	-0.1	-0.4	-0.5	-0.4	0.0	0.0	-0.4	-0.2	0.0	-0.1	0.0	0.0	-0.1	-1.7	-2.5	-0.4	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
DCs	MAG_{POP}	3.9	3.0	3.4	2.7	0.0	0.0	0.1	3.1	0.9	0.0	0.1	0.3	1.3	0.0	0.7	0.7	0.0	1.3	0.0	0.0	0.0	0.0	0.0	0.0
	MAG_{REF}	4.1	2.5	3.0	0.0	0.0	0.0	0.6	0.0	0.0	0.1	0.0	0.0	0.0	1.9	2.0	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	IQR_{POP}	1.1	1.6	1.0	2.0	0.5	0.5	0.5	3.9	1.7	0.5	0.5	1.5	3.9	0.5	1.4	1.2	0.5	2.3	0.5	0.5	0.5	0.5	0.5	0.5
	IQR_{REF}	0.7	1.3	1.3	0.5	0.5	0.5	1.5	0.5	0.5	0.7	0.5	0.5	0.5	2.4	3.1	1.0	0.6	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	MEM	0.2	0.5	0.8	2.6	0.0	0.0	-3.2	3.0	0.2	-0.5	0.1	-0.5	0.5	-7.6	-3.4	-0.1	-0.4	0.7	0.0	0.0	0.0	0.0	0.0	0.0
	MAG_{DIFF-S}	-0.4	1.1	0.7	5.6	0	0	-1	6.6	1.8	-0.1	0.1	0.6	2.6	-3.9	-2.8	0.2	0	2.6	0	0	0	0	0	0
	MAG_{DIFF}	-0.2	0.5	0.4	2.7	0.0	0.0	-0.5	3.1	0.9	-0.1	0.1	0.3	1.3	-1.9	-1.3	0.1	0.0	1.3	0.0	0.0	0.0	0.0	0.0	0.0
	IQR_{DIFF}	0.4	0.3	-0.2	1.5	0.0	0.0	-1.0	3.4	1.2	-0.2	0.0	1.0	3.4	-1.9	-1.7	0.2	-0.1	1.8	0.0	0.0	0.0	0.0	0.0	0.0
IgM- B cells	MAG_{POP}	4.4	3.5	3.1	3.1	0.0	2.5	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0
	MAG_{REF}	4.1	2.5	3.0	0.0	0.0	0.0	0.6	0.0	0.0	0.1	0.0	0.0	0.0	1.9	2.0	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	IQR_{POP}	0.8	1.4	1.0	1.3	0.5	0.7	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1.0	0.5	0.5	0.5	0.6	0.5	0.5	0.5	0.5
	IQR_{REF}	0.7	1.3	1.3	0.5	0.5	0.5	1.5	0.5	0.5	0.6	0.5	0.5	0.5	2.4	3.0	1.0	0.6	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	MEM	0.2	1.3	0.5	3.3	0.0	3.0	-3.3	0.0	0.1	-0.5	0.0	0.0	-0.1	-7.6	-9.4	0.1	-0.4	0.0	0.0	0.1	0.0	0.0	0.0	0.0
	MAG_{DIFF-S}	0.4	2.1	0.2	6.5	0	5.2	-1.1	0	0.1	-0.2	0	0	0	-3.9	-4.2	0.2	0	0	0	0.3	0	0	0	0
	MAG_{DIFF}	0.2	1.0	0.1	3.1	0.0	2.5	-0.5	0.0	0.1	-0.1	0.0	0.0	0.0	-1.9	-2.0	0.1	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0
	IQR_{DIFF}	0.0	0.0	-0.3	0.8	0.0	0.2	-1.0	0.0	0.0	-0.1	0.0	0.0	0.0	-1.9	-2.5	0.0	-0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0
IgM+ B cells	MAG_{POP}	4.3	2.6	2.7	3.7	3.0	2.5	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0
	MAG_{REF}	4.1	2.5	3.0	0.0	0.0	0.0	0.6	0.0	0.0	0.1	0.0	0.0	0.0	2.0	2.3	0.6	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	IQR_{POP}	0.7	1.2	0.9	1.4	2.6	0.7	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1.1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	IQR_{REF}	0.7	1.4	1.3	0.5	0.5	0.5	1.6	0.5	0.5	0.7	0.5	0.5	0.6	2.4	3.0	0.9	0.7	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	MEM	0.4	0.5	-1.0	4.0	2.9	3.0	-3.6	0.0	0.1	-0.6	0.0	0.0	-0.3	-7.5	-9.6	0.2	-0.6	0.0	0.0	0.1	0.0	0.0	0.0	0.0
	MAG_{DIFF-S}	0.4	0.3	-0.7	7.7	6.3	5.2	-1.2	0	0.2	-0.2	0	0	0	-4	-4.8	0.5	0	0	0	0.2	0	0	0	0
	MAG_{DIFF}	0.2	0.2	-0.3	3.7	3.0	2.5	-0.6	0.0	0.1	-0.1	0.0	0.0	0.0	-1.9	-2.3	0.3	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0
	IQR_{DIFF}	-0.1	-0.2	-0.4	0.9	2.1	0.2	-1.1	0.0	0.0	-0.2	0.0	0.0	-0.1	-1.9	-2.5	0.1	-0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Monocytes	MAG_{POP}	4.1	2.3	4.2	2.1	0.0	0.0	0.2	3.6	3.5	2.7	2.9	2.9	0.1	0.0	1.1	1.3	0.2	0.4	0.1	0.0	0.0	0.1	0.0	0.0
	MAG_{REF}	4.2	2.5	2.9	0.0	0.0	0.0	0.6	0.0	0.0	0.1	0.0	0.0	0.0	2.0	2.4	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	IQR_{POP}	0.7	1.1	0.7	1.5	0.5	0.5	0.5	0.8	0.6	0.7	0.7	1.2	0.5	0.5	0.9	0.8	0.6	0.8	0.5	0.5	0.5	0.5	0.5	0.5
	IQR_{REF}	0.7	1.4	1.2	0.5	0.5	0.5	1.7	0.5	0.5	0.5	0.5	0.5	0.6	2.2	3.1	0.9	0.6	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	MEM	-0.2	-0.5	2.6	1.9	0.0	0.0	-3.5	4.3	4.5	3.1	3.5	3.1	0.2	-7.2	-5.0	1.1	0.3	0.0	0.1	0.0	0.0	0.1	0.0	0.0
	MAG_{DIFF-S}	-0.2	-0.3	2.6	4.4	0	0	-0.9	7.5	7.4	5.5	6	6.1	0	-4.2	-2.7	1.5	0.4	0.7	0.1	0	0	0.1	0	0
	MAG_{DIFF}	-0.1	-0.2	1.3	2.1	0.0	0.0	-0.5	3.6	3.5	2.6	2.9	2.9	0.0	-2.0	-1.3	0.8	0.2	0.4	0.1	0.0	0.0	0.1	0.0	0.0
	IQR_{DIFF}	0.0	-0.2	-0.5	1.0	0.0	0.0	-1.1	0.3	0.1	0.2	0.2	0.7	-0.1	-1.7	-2.2	-0.1	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0
NK cells	MAG_{POP}	3.8	2.6	1.6	0.0	0.0	0.0	0.5	0.8	0.0	0.6	0.0	0.0	4.3	0.0	0.0	1.1	0.1	0.0	0.8	0.0	0.0	0.0	0.0	0.0
	MAG_{REF}	4.2	2.4	3.2	0.0	0.0	0.0	0.5	0.0	0.0	0.1	0.0	0.0	0.0	2.1	2.6	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	IQR_{POP}	0.7	0.9	0.9	0.5	0.5	0.5	2.4	1.6	0.5	0.9	0.5	0.5	0.8	0.5	0.5	1.3	0.5	0.5	1.0	0.5	0.5	0.5	0.5	0.5
	IQR_{REF}	0.7	1.5	1.0	1.0	0.5	0.5	1.3	0.5	0.5	0.5	0.5	0.5	0.5	1.6	2.8	0.9	0.7	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	MEM	-0.6	1.0	-2.3	1.2	0.0	0.0	0.6	0.1	0.0	0.1	0.0	0.0	5.2	-5.6	-9.6	0.3	0.5	0.0	0.4	0.0	0.0	0.0	0.0	0.0
	MAG_{DIFF-S}	-0.9	0.2	-3.3	0	0	0	0	1.6	0	1	0	0	9.1	-4.3	-5.3	1.1	0	0	1.7	0	0	0	0	0
	MAG_{DIFF}	-0.4	0.1	-1.6	0.0	0.0	0.0	0.0	0.8	0.0	0.5	0.0	0.0	4.3	-2.1	-2.5	0.5	0.0	0.0	0.8	0.0	0.0	0.0	0.0	0.0
	IQR_{DIFF}	0.0	-0.6	-0.1	-0.5	0.0	0.0	1.1	1.1	0.0	0.4	0.0	0.0	0.3	-1.1	-2.3	0.4	-0.2	0.0	0.5	0.0	0.0	0.0	0.0	0.0

The MEM scores and magnitude difference values were most divergent for features that were negatively enriched on specific cell populations (Figure 4-3, Table 4-2). For example, the MEM score of CD8 on CD4+ T cells ($CD4^+$ T cells: $CD8^{-7}$) was lower than the difference in median CD8 expression between CD4+ T cells and the rest of the cells in the sample ($CD4^+$ T cells: MAG_{DIFF} CD8 = -0.3). MEM therefore emphasized the negative enrichment (i.e. lack of expression) of CD8 on these cells. Similarly, the MEM score for CD4 on most other populations (e.g. $CD8^-$ T cells: $CD4^{-10}$) was lower than that population's median difference (e.g. $CD8^+$ T cells MAG_{DIFF} CD4 = -2.0). Thus, MEM correctly emphasizes the negative enrichment of CD4 on cells other than $CD4^+$ T cells and APCs (Table 4-2). Incorporating information about feature variability therefore allows MEM to capture negative enrichment that is not reflected in magnitude difference alone (Figure 4-3 and Table 4-2).

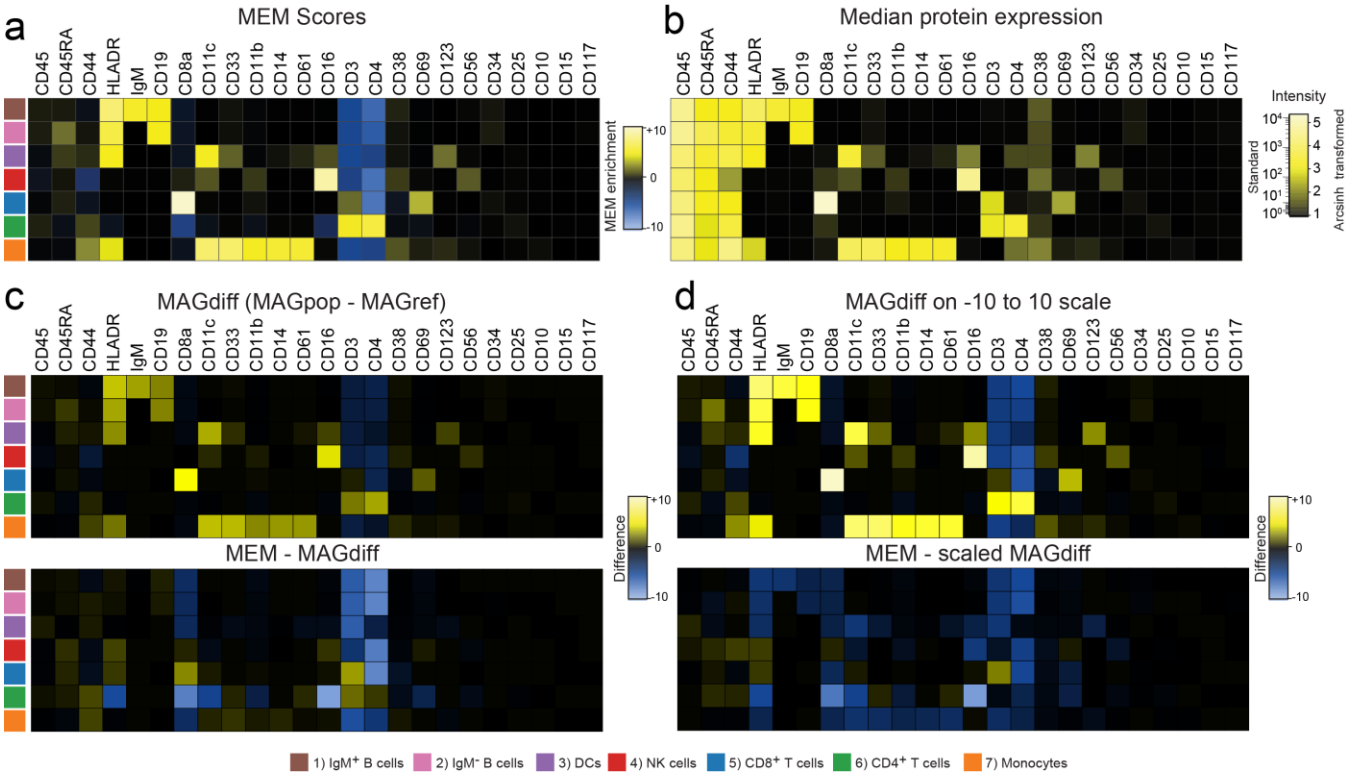

Figure 4-3. MEM captures negative feature enrichment that is not apparent in the difference between population medians. For 7 major normal human PBMC cell populations and 25 proteins, heatmaps show a) MEM scores; b) median expression levels (MAGpop); c) the difference between median expression value of the population and the median expression of the other immune cells (MAGdiff) (top) and the difference between the MEM score and the MAGdiff value to highlight differences between these two methods of population comparison (bottom); and d) the same comparisons made in c) using scaled MAGdiff values to -10 to +10 scale (linearly; x/max(MAGdiff)*10). Scaling is meant to allow for a more direct comparison to the MEM scores.

To test the hypothesis that the features with the highest MEM scores would be important for computational cluster formation, the 25 proteins measured in Dataset A (Figure 4-2b) were sorted in six ways:

1) high to low MEM score, 2) high to low median value, 3) high to low MAG_{DIFF}, 4) high to low z-score, 5) high to low K-S statistic, and 6) randomly (Table 4-3). Z-score and K-S statistic values are shown in Table 4-4. The proteins were then sequentially, cumulatively excluded from use in k-means clustering and F-measure was used to measure clustering accuracy (Fig. 4-2c and Figure 4-4). The order in which markers were excluded is shown in Table 4-3. Random exclusion was performed 15 times and the average result is shown (Fig. 4-2c). Clustering accuracy was most impacted by excluding proteins based on MEM score. F-measure dropped to 0.75 after removing the proteins with the top 6 MEM scores, whereas this drop in f-measure was only seen after removing the 14 highest markers based on MAG_{DIFF}, the 13 highest markers based on z-score, and the 12 highest markers based on K-S statistic values (Figure 4-4). Removing markers based on median value was not significantly different from removing markers randomly until after the 15 markers with the highest median signal intensity were excluded (Figure 4-4). The same analysis was also performed with viSNE in place of k-means clustering to visualize the loss of population resolution when markers were excluded based on MEM or median scores (Fig. 4-5c). In this case, loss of accuracy as markers were excluded by MEM values was reflected in the viSNE map as a loss of well-separated population islands comprised of one cell type. These results indicated that MEM enrichment scores captured the markers that were important to cell identity better than traditional comparisons based solely on median protein expression.

Table 4-3. K-S and z-score values for immune cell populations in Fig. 1

		CD19	CD117	CD11b	CD4	CD8	CD20	CD34	CD61	CD123	CD45RA	CD45	CD10	CD33	CD11c	CD14	CD69	CD15	CD16	CD44	CD38	CD25	CD3	IgM	HLADR	CD56
K-S	CD4+ T	1.8	1.2	3.3	9.7	4.1	3.8	2.7	2.2	1.8	2.0	2.0	0.1	2.3	3.8	2.5	3.7	1.3	3.4	3.2	1.7	3.6	6.7	1.2	3.7	3.6
	CD8+ T	1.0	1.2	1.3	6.2	10.0	6.5	3.5	1.1	1.0	1.4	1.0	0.7	1.5	2.3	1.2	8.3	0.1	1.9	2.8	3.3	2.1	4.1	0.8	2.0	0.3
	DC	0.8	0.7	4.5	4.0	3.1	1.5	0.3	3.1	5.8	2.6	1.8	0.2	4.7	4.8	2.8	1.8	0.6	3.4	1.8	0.9	2.1	6.7	0.9	7.8	0.8
	IgM- B	9.3	3.6	1.7	5.9	4.8	7.6	3.4	0.7	0.4	4.3	1.6	0.4	2.2	1.0	0.9	3.0	0.6	1.7	0.9	0.7	0.5	6.6	1.5	7.9	1.1
	IgM+ B	9.8	3.6	2.9	6.4	5.4	8.4	3.5	0.8	0.9	1.3	1.5	0.5	2.6	1.7	1.2	3.1	0.4	1.9	2.1	1.4	0.7	6.8	9.4	8.5	0.9
	Mono	0.9	0.5	9.8	4.9	3.0	2.5	1.2	9.2	5.6	1.0	0.8	1.5	9.9	9.4	9.7	2.7	1.4	1.7	6.6	4.2	2.5	7.3	0.6	8.2	1.8
	NK	1.8	1.4	3.6	6.6	1.9	2.0	1.6	1.1	0.9	1.5	3.2	0.6	1.3	4.7	1.1	1.0	1.8	9.8	7.3	2.7	2.8	7.7	2.0	3.2	6.1
z-score	CD4+ T	-0.6	-0.3	-0.9	4.6	-0.7	-1.0	-0.6	-0.6	-0.5	-0.8	0.6	0.0	-0.7	-1.0	-0.7	-1.0	-0.3	-0.9	0.7	-0.4	1.6	1.9	-0.4	-0.9	-0.9
	CD8+ T	-0.3	0.3	-0.1	-1.5	5.6	1.4	0.7	-0.4	-0.3	0.4	-0.2	-0.2	-0.5	-0.7	-0.5	4.2	0.0	-0.4	-0.4	-1.0	-0.5	0.9	-0.3	-0.5	0.1
	DC	-0.1	-0.2	-0.9	-0.6	-0.8	-0.2	0.1	0.7	4.4	0.6	-0.4	0.0	1.1	1.9	0.2	-0.5	0.1	1.0	0.4	0.3	-0.5	-1.7	0.2	2.3	0.0
	IgM- B	4.4	1.2	-0.5	-1.6	-1.0	3.1	1.2	-0.2	0.0	1.2	0.4	0.1	0.0	-0.3	-0.3	-0.7	-0.1	-0.6	0.2	0.3	0.0	-1.6	-0.3	2.6	-0.3
	IgM+ B	7.6	1.4	-0.7	-1.7	-1.1	3.9	1.2	-0.3	0.1	0.5	0.5	0.2	0.0	-0.5	-0.4	-0.8	-0.1	-0.6	-0.3	0.4	-0.2	-1.7	9.9	3.8	-0.2
	Mono	-0.3	0.1	5.1	-0.3	-0.8	-0.7	0.3	6.2	1.9	0.0	-0.2	0.5	8.7	5.1	10.0	0.1	0.4	-0.4	1.8	1.1	-0.6	-1.9	-0.3	2.3	0.0
	NK	-0.4	-0.1	0.5	-1.9	-0.2	-0.6	-0.4	-0.4	-0.3	0.3	-1.0	-0.1	-0.5	0.8	-0.2	-0.2	0.5	7.4	-2.8	0.9	-0.7	-2.2	-0.5	-0.8	2.5

Each population (POP) was compared to all non-population cells (REF). K-S statistic was calculated as a comparison between POP and REF distributions for each protein. Z-score was calculated as the number of standard deviations the POP mean was from the REF mean. K-S and z-score were signed and scaled to -10 to +10 as for MEM values.

Table 4-4. Order of marker exclusion for clustering and f-measure (high to low)

MEM	MAG	MAG_{DIFF}	z-score	K-S
CD4	CD8a	CD8a	CD8	CD14
CD16	CD45	CD16	CD33	IgM
CD3	CD16	CD11c	CD16	CD33
CD8a	CD44	HLA-DR	CD11b	CD19
CD11c	CD11c	CD33	CD19	CD16
HLA-DR	HLA-DR	CD4	CD14	CD61
CD33	CD33	CD61	CD4	CD8
CD61	CD4	CD14	CD11c	CD11b
CD14	CD61	CD19	IgM	CD11c
CD19	CD45RA	CD11b	CD61	CD4
CD11b	CD14	IgM	HLADR	CD123
CD69	CD11b	CD3	CD20	CD69
CD44	CD19	CD20	CD69	CD20
CD20	IgM	CD69	CD3	HLADR
IgM	CD3	CD44	CD44	CD44
CD38	CD20	CD56	CD56	CD56
CD45RA	CD69	CD38	CD123	CD3
CD56	CD38	CD123	CD45RA	CD25
CD45	CD56	CD45	CD38	CD117
CD34	CD123	CD45RA	CD25	CD34
CD10	CD34	CD34	CD117	CD45RA
CD117	CD25	CD25	CD34	CD38

CD123	CD10	CD10	CD45	CD45
CD25	CD117	CD15	CD15	CD15
CD15	CD15	CD117	CD10	CD10

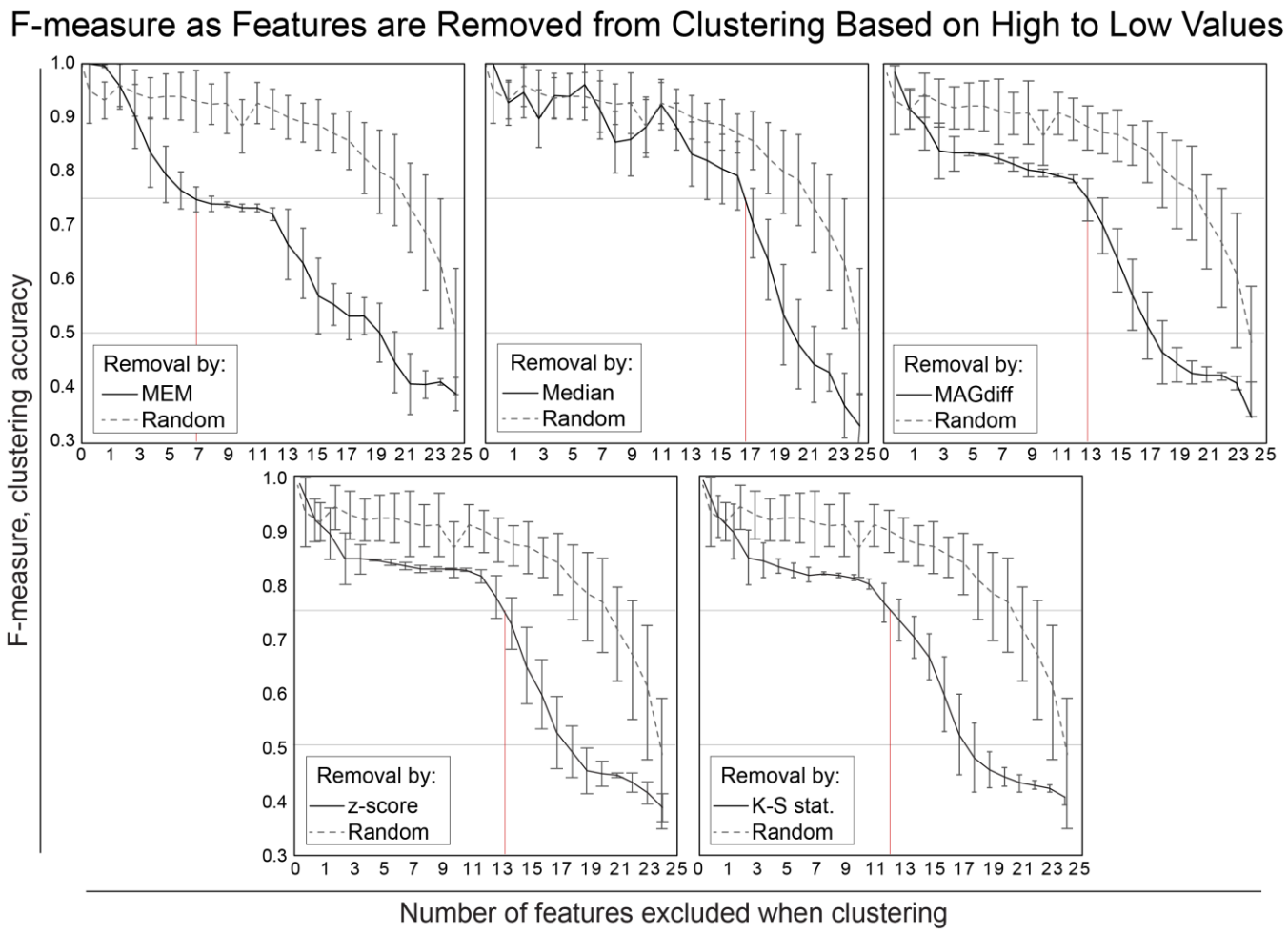

Figure 4-4. MEM highly scores markers that are important to clustering accuracy. Markers were sequentially and cumulatively excluded from k-means cluster analysis of Dataset A, from high to low, sorted based on 5 different statistics or scores (marker order shown in Table 4-3): MEM, median, median difference (MAG_{DIFF}), z-score, and Kolmogorov-Smirnov (K-S) statistic. Clustering accuracy was quantified as the f-measure where true cluster identity was assumed to be the clusters formed by clustering on all 25 markers in the dataset. The moving average of the f-measure is shown. Error bars represent the standard error. The vertical red line indicates the number of excluded features at which the f-measure reached 0.75.

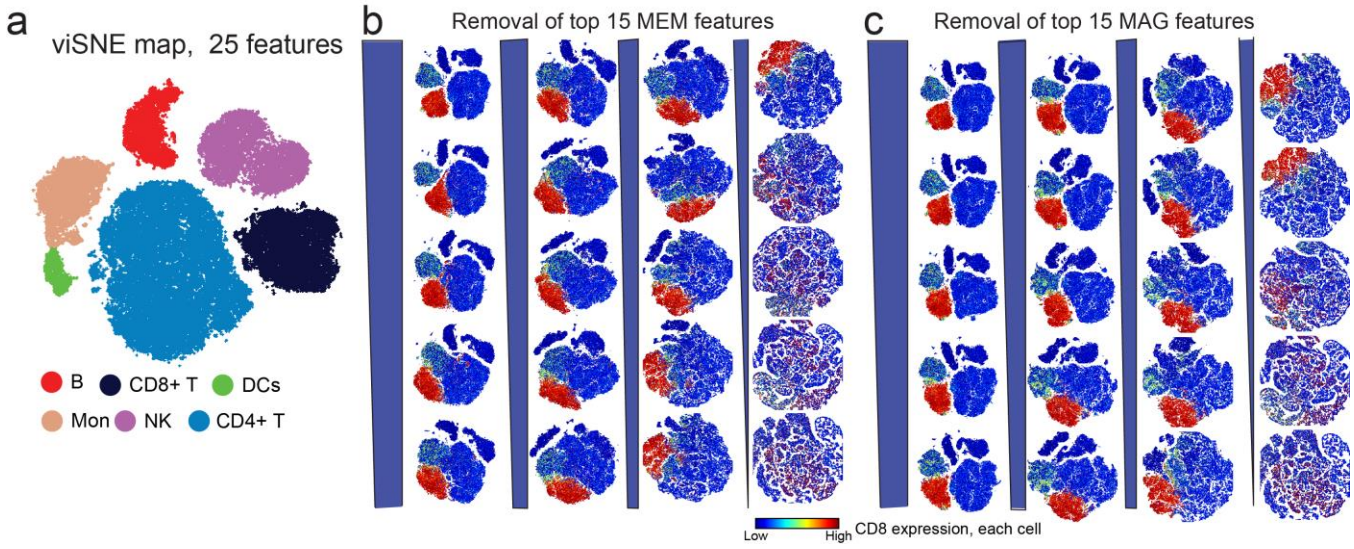

Figure 4-5. MEM highly scores markers that are important to viSNE mapping a) viSNE map for healthy human blood, built using 25 surface protein markers. Populations were identified by expert analysis and color coded. b). Top to bottom, left to right: viSNE maps generated as markers were iteratively, cumulatively excluded based on their MEM scores (high to low absolute value). Heat intensity for each cell indicates CD8 expression. c) Top to bottom, left to right: viSNE maps generated as markers were iteratively, cumulatively excluded based on their median scores (high to low absolute value). Heat intensity for each cell indicates CD8 expression.

Dataset B was mass cytometry data quantifying expression of 31 proteins on healthy human bone marrow(Bendall, Simonds et al. 2011). The diverse cell types identified in this dataset included mature cells, such as $CD4^+$ T cells recirculating from blood, and cells developing along a continuum originating from hematopoietic stem cells (HSCs), such as myeloid and lymphoid cells. Computational and expert analysis by the authors had previously identified 23 populations of cells(Bendall, Simonds et al. 2011) that were analyzed here by MEM (Figure 4-6a). MEM effectively quantified enrichment of proteins marking stem and progenitor cells as well as mature cells. For example, the cell subset labeled by Bendall et al. as HSCs was highly enriched for CD34 ($CD34^{+6}$) and negatively enriched for CD45 ($CD45^{-5}$), consistent with the authors' original interpretation and longstanding biological observations(Civin, Strauss et al. 1984, Doulatov, Notta et al. 2012). As cells matured, MEM enrichment scores for HLA-DR, CD19, and CD3, three proteins functionally involved in shaping cell identity, increased on respective subsets of APCs, B cells, and T cells (Figure 4-6). Notably, MEM scores for CD45 expression in Dataset B varied across subsets and correlated with differentiation and maturation (Figure 4-6). This result in bone marrow contrasted with the results in Dataset A, where MEM scores for CD45 were uniformly near zero because the mature blood cell subsets all expressed high levels of CD45 (Figure 4-2).

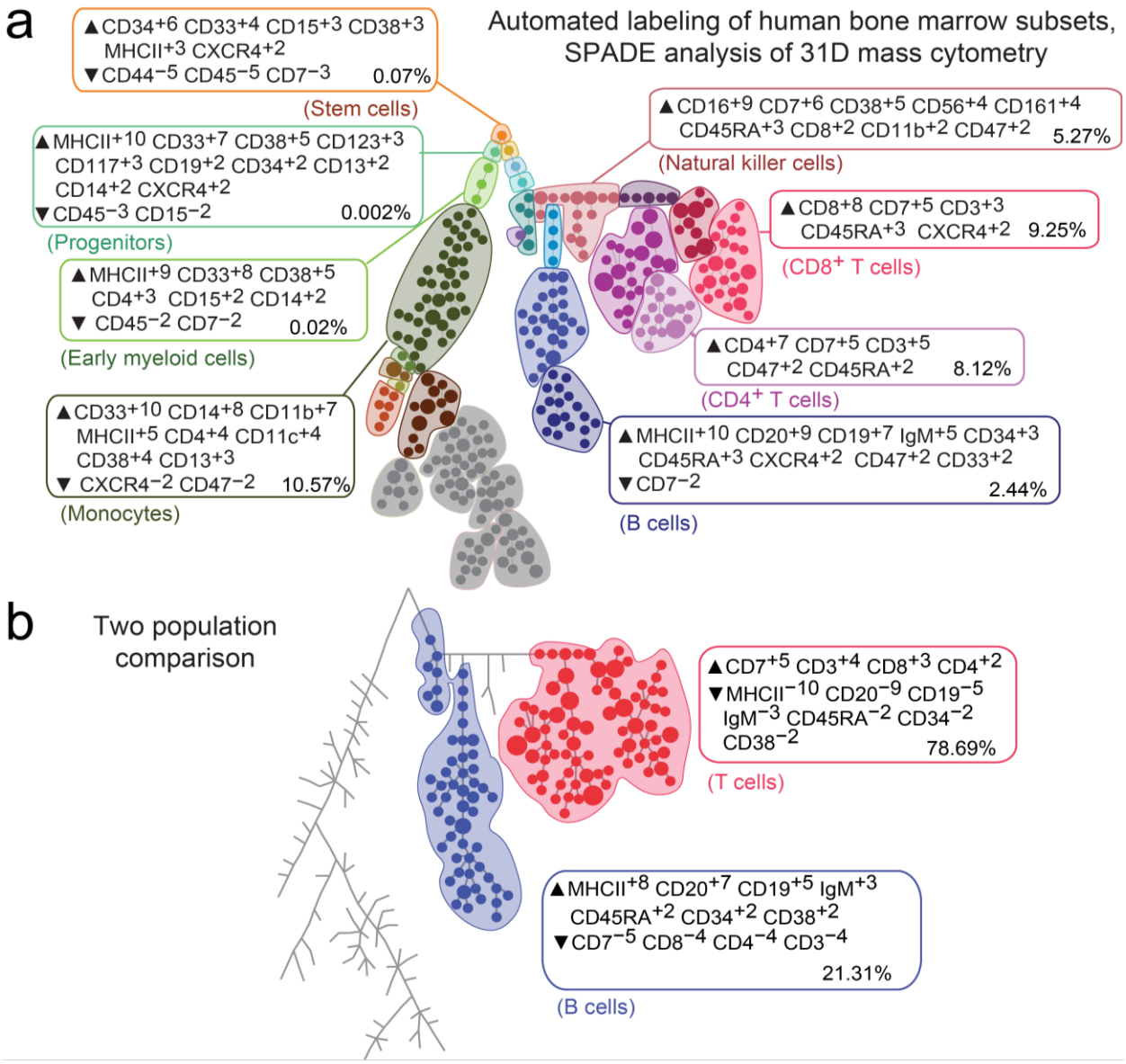

Figure 4-6: MEM labels report enriched protein expression in human bone marrow cells in Dataset B. a) Cells from normal human bone marrow were previously grouped into 23 nucleated cell populations using spanning-tree progression analysis of density normalized events (SPADE) and expert review of 31D mass cytometry data(Bendall, Simonds et al. 2011). MEM labels are shown for 8 cell subsets that ranged in frequency from 0.002% to 10.57% of total cells. All non-population cells were used as reference for MEM. Previously determined expert labels are included in parenthesis for comparison to automatically generated MEM labels. Additional MEM analysis of this dataset is available for all 28 identified populations (Figure 4-14) and using stem cells as a common comparison point (Figure 4-5). b) Pairwise analysis of the same dataset compared just bone marrow T cells (red) and B cells (blue). The percentage of each population in the comparison is indicated in the lower right of the MEM label.

To assess how MEM scores captured intra-population variability, closely related populations from human bone marrow were grouped and new MEM labels calculated in a pairwise comparison of $CD19^+$ B cells and $CD3^+$ T cells (Dataset B, Fig. 4-6b). As expected, $CD3^+$ T cells had a lower MEM score for CD8 than the subset of

$CD3^+$ $CD8^+$ T cells ($CD8^{+3}$ vs. $CD8^{+8}$, respectively) because $CD8^-$ $CD4^+$ T cells were part of the parental $CD3^+$ T cell population. Likewise, IgM enrichment was lower in all $CD19^+$ B cells than in the IgM^+ subset of $CD19^+$ B cells (IgM^{+3} vs. IgM^{+6}). As a general rule, MEM scores will approach median values as feature variability within populations decreases (Figure 4-7). Thus, MEM captures feature enrichment and heterogeneity better than median in diverse populations, as in Fig. 4-2c.

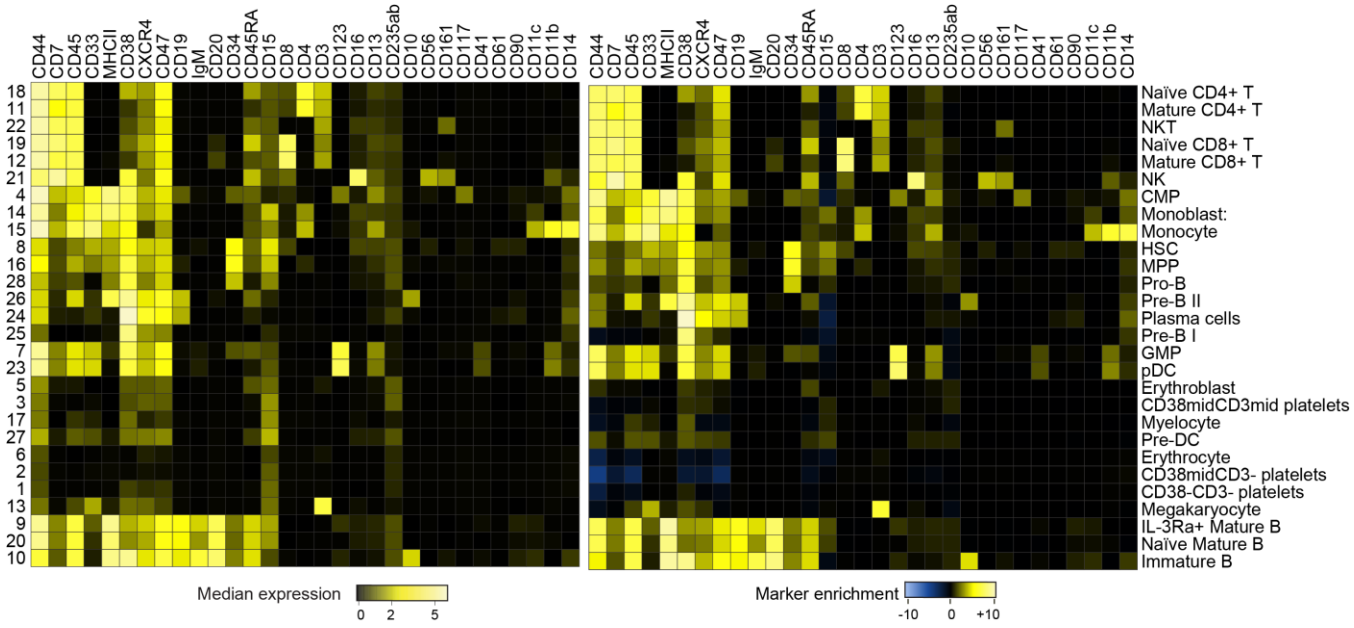

Figure 4-7. MEM scores largely reflect median expression values for relatively homogenous populations. Heatmaps show median intensity of protein expression (left) and protein enrichment by MEM (right) for measured proteins in 28 populations characterized as relatively homogeneous for established cell types by expert analysis (rows). Each population was compared to the other 27 subsets for the MEM analysis. MEM scores approach median expression values in homogeneous populations because the contribution of variance approaches zero.

Dataset C was mass cytometry data quantifying expression of 38 proteins on murine cells from eight tissues(Becher, Schlitzer et al. 2014). This dataset was chosen to evaluate MEM population comparisons across tissues in a non-human model and to test the hypothesis that MEM would objectively identify the most enriched features of novel cell subsets. Dataset C was analyzed three different ways by MEM: 1) *comprehensive MEM phenotyping* highlighted the major features of each population, as in Becher et al. (Figure 4-8); 2) *subset-focused MEM phenotyping* identified the distinguishing features of novel cell subsets identified by Becher et al., such as innate lymphoid cell (ILC) population 28 (Figure 4-9 and Figure 4-10); 3) *tissue-specific MEM phenotyping* provided a new way of highlighting how neutrophil phenotype varied across tissues. While comprehensive

phenotyping provided an overview of the major differences between cell subsets (Figure 4-8), subset-focused phenotyping better captured the features experts had identified as uniquely enriched on novel cell populations (Figure 4-9). For example, cluster 28 was a novel population identified by Becher et al. as CD11b^{int} NK cells(Becher, Schlitzer et al. 2014). The MEM label for cluster 28 within ILCs was ▲CD11b^{+5} CD62L^{+3} ▼CD4^{-7} CD103^{-4} Terr119^{-3} (Figure 4-9b and Figure 4-10). This MEM label captured the key feature of this novel innate lymphoid cell subset (CD11b^{int}) and highlighted additional features that can be used to match this subset to cells identified by others (i.e., to cytotype the population).

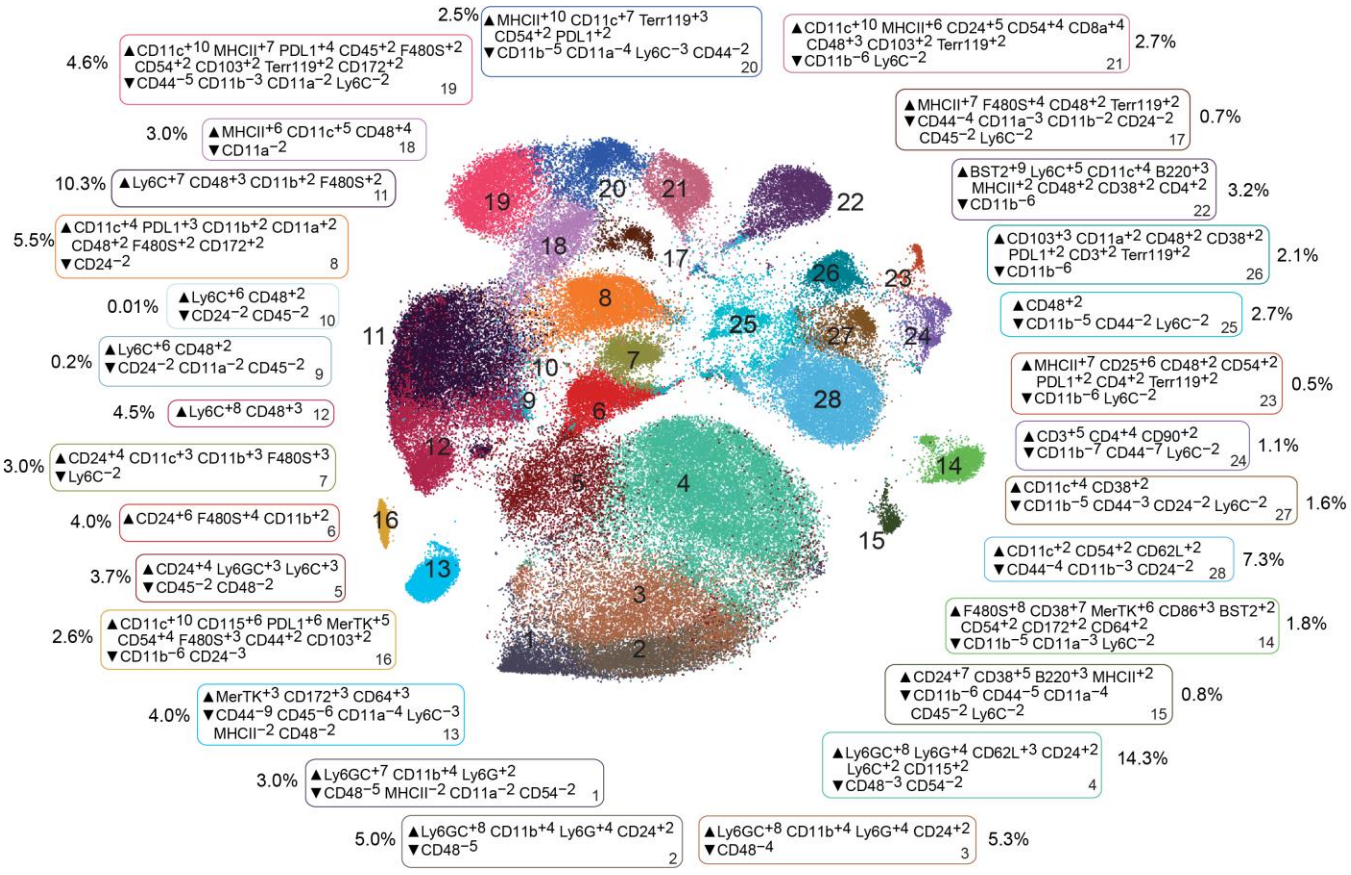

Figure 4-8. MEM quantifies feature enrichment on 28 automatically identified murine myeloid cell subsets. MEM labels are shown for each of 28 murine myeloid populations automatically identified by Becher, et al(Becher, Schlitzer et al. 2014). Each population was compared to the remaining 27 populations as reference for the MEM calculation.

To compare tissue specific phenotypes, MEM was next used in a novel analysis characterizing neutrophils from each of the 8 tissues (Figure 4-9c). CD54 (ICAM-1), was highly enriched on lung neutrophils (CD54^{+8}) as opposed to on neutrophils from other tissues, such as bone marrow (CD54^{-3}). CD54 plays a

crucial role in neutrophil recruitment to the lungs(Basit, Reutershan et al. 2006). MEM also revealed that CD62L, a molecule that is shed from neutrophils that have recently left the bone marrow, was most enriched on bone marrow neutrophils(Furze and Rankin 2008). Other markers, including CD24, HLA-DR, and CD48, also displayed tissue-specific enrichment on neutrophils. Therefore, changing the context of the MEM comparison and using alternative reference populations can reveal enrichment of biologically relevant features that is not apparent with standard approaches, such as median analysis. MEM comparisons can also capture the environment in which a subset is normally found, and this was observed to provide additional information about the tissue context that might be especially helpful in translational applications of MEM in disease settings. These results indicate that MEM labels complement unbiased population discovery and effectively characterize *cyto incognito*(Irish 2014) by providing unbiased descriptions that correctly capture key features of novel cell types.

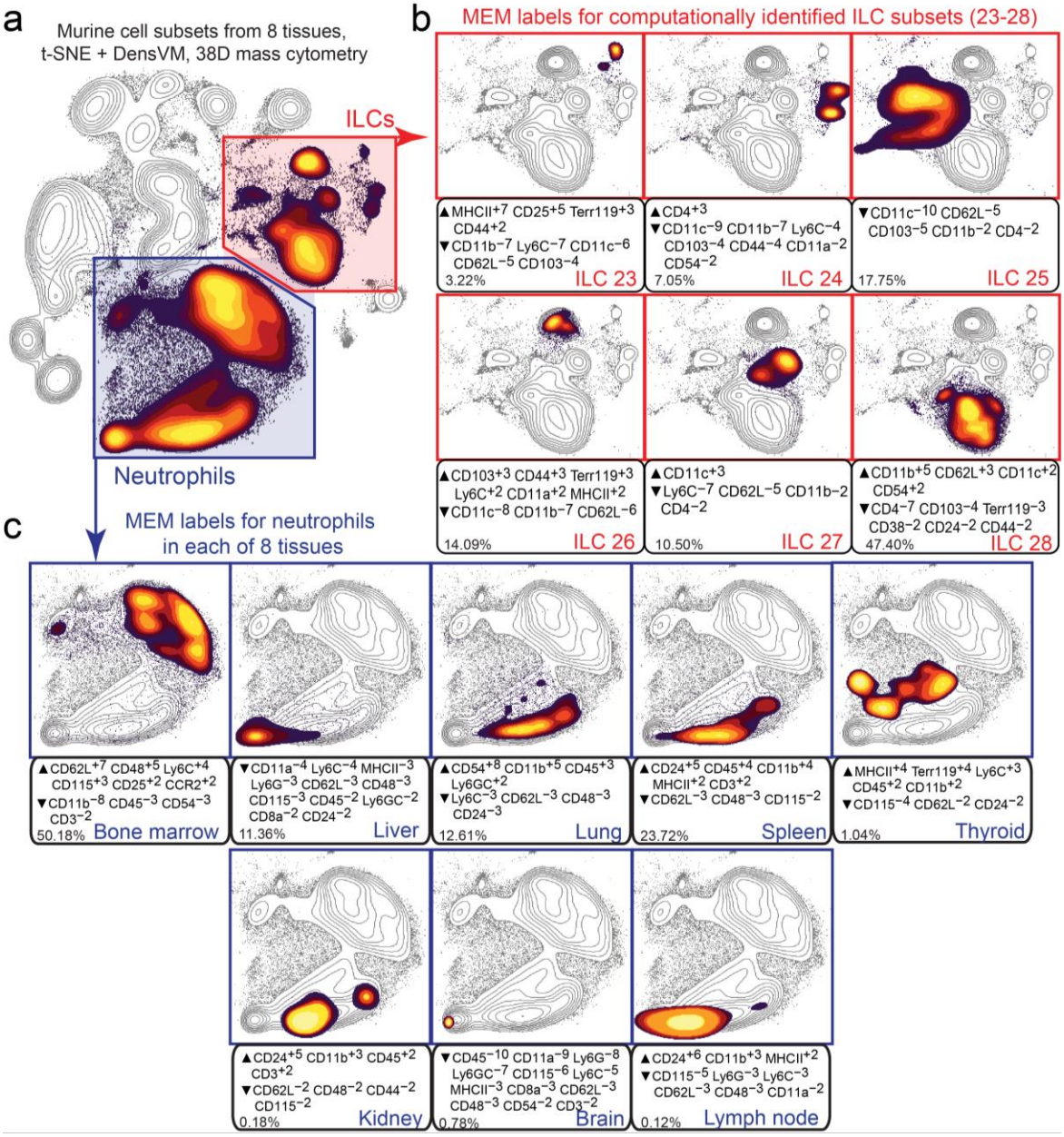

Figure 4-9. MEM labels characterize subset specific and tissue specific phenotypes of murine innate lymphoid cells and neutrophils in Dataset C. a) Previous computational analysis automatically grouped murine cells into 28 populations using t-SNE and DensVM analysis of 38 proteins measured by mass cytometry(Becher, Schlitzer et al. 2014). b) Density of each of 6 computationally defined populations of ILCs is overlaid in color on a greyscale contour plot of cell density for all ILCs (red region from a). MEM labels characterized the enriched features of each computationally identified population, such as ILC subset 28, which was distinguished as $CD11b^{+5}$. c) MEM labels were generated for all neutrophil subsets according to tissue site (labeled in blue). Density of each neutrophil population is overlaid in color on a contour t-SNE plot of all neutrophils (blue region in a). The percentage of each population in the comparison is indicated in the lower left of the MEM label. Additional MEM analysis of this dataset is available for all 28 identified populations (Figure 4-8).

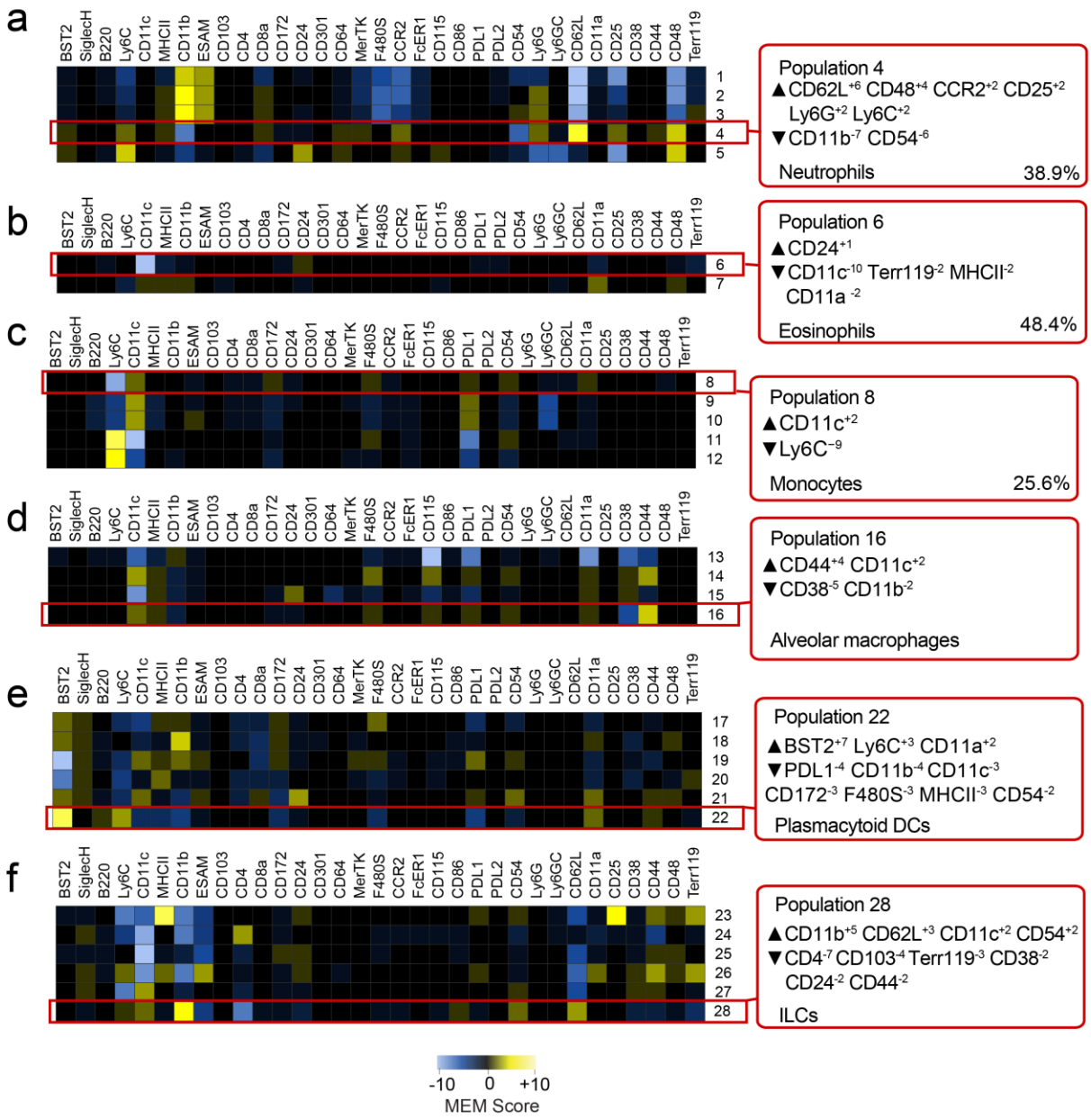

Figure 4-10. Focused MEM analysis quantifies feature enrichment within phenotypically similar groups of cells a-f) Focused MEM analysis on murine myeloid cell subsets. A MEM label for one population within each group is shown as an example. Groups were defined as the 6 major murine subgroups identified by t-SNE and DensVM by Becher et al.(Becher, Schlitzer et al. 2014).

An important aspect of MEM is generation of machine-readable quantitative labels that can be used to register population identities across samples and studies. A MEM label for a newly discovered population can be compared quantitatively against a reference set of established MEM labels or a MEM label reported in a paper. To illustrate this idea, pairwise similarity of MEM scores was calculated for 80 populations of cells from 7 different studies including healthy $CD4^+$ T cell and B cell (Figure 4-11). Cells had highly similar MEM scores within each major cell type, regardless of platform (mass vs fluorescence flow cytometry), study, or tissue source. For example, T cells run on mass cytometry from different blood donors were 97% ± 1.3 similar to each other, 85% ± 1.9 similar to T cells from blood run on fluorescence flow cytometry, and 87% ± 2.1 similar to T cells from tonsil run on mass cytometry (Figure 4-11). However, these cells were 66.9% ± 13 similar to any B cell population. This indicates that MEM scores provide a way to communicate cell identity and to quantify similarities of cell types from the text label alone.

Dataset D included 52 populations of tumor infiltrating APCs, tumor infiltrating T cells, and non-immune malignant tumor cells identified in human glioma tumors(Leelatian, Doxie et al. 2016). To obtain these populations, each tumor was analyzed by viSNE and cell subsets were expert gated solely on t-SNE cluster density (Figure 4-12). To determine whether MEM could distinguish immune cell subsets from other tumor cell types with limited information, MEM scores were calculated using only 9 markers (S100B, TUJ1, GFAP, Nestin, MET, PDGFRα,EGFR, HLA-DR, and CD44) that were expected to be expressed on cancer cells (Figure 4-13a). The 52 populations were grouped into 13 major groups of cell types based on MEM enrichment of the 9 analyzed proteins and these groups interpreted as either tumor infiltrating APCs (Figure 4-13b, blue), tumor infiltrating T cells (Figure 4-13b, green), or non-immune tumor cells (Figure 4-13b, red). To confirm cell identity, four protein features that had been excluded from MEM analysis were assessed (Figure 4-13c, CD45, CD3, CD45RO, and CD64). CD45 and CD3 were used to confirm T cell identity and CD45 and CD64 were used to confirm APC identity. MEM correctly identified both immune cell subsets from all tumor types without using key immune lineage markers and without using healthy populations (e.g. APCs from blood or tonsil) to guide the clustering. Thus, MEM labels distinguished populations of cells based on non-traditional features and in a disease context.

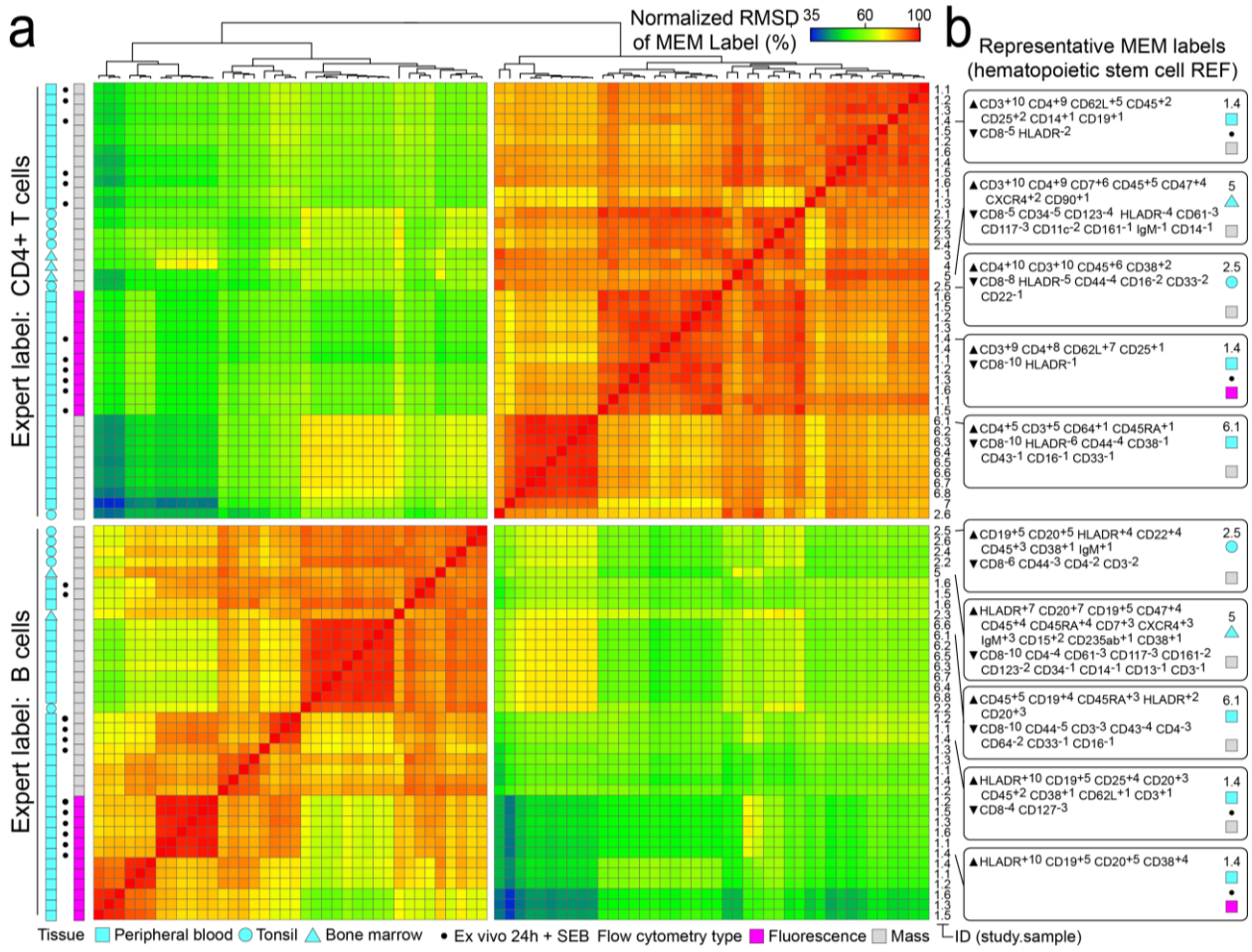

Figure 4-11. Classification based solely on MEM label groups samples of T cells and B cells measured in diverse studies using different cytometry platforms. A) MEM label values were compared for each of 80 populations (CD4+ T cells and B cells) from 3 human tissues representing 6 mass cytometry studies and 1 fluorescence flow cytometry study. The similarity score for two populations was 100% when MEM label exponents were identical for all of the shared proteins. Populations are shown clustered according to MEM label percent similarity. Tissue type, source study, and individual sample IDs are indicated to the right. Published study references: Study references: 1) Nicholas KJ, et al., Cytometry A 2016(Nicholas, Greenplate et al. 2016); 2) Polikowsky et al., J Immunol. 2015(Polikowsky, Wogsland et al. 2015); 3) Ferrell et al., PLOS One 2016(Ferrell, Diggins et al. 2016); 4) Amir et al., Nature Biotech 2013(Amir el, Davis et al. 2013); 5) Bendall et al., Science 2011(Bendall, Simonds et al. 2011); 6) unpublished data courtesy of Greenplate AR, Irish lab; 7) Leelatian et al., Methods Mol Bio 2015(Leelatian, Diggins et al. 2015). *indicates samples stimulated by bacterial superantigen *Staphlococcus* enterotoxin B(Nicholas, Greenplate et al. 2016). B) Representative MEM labels for CD4+ T cells (top) and B cells (bottom) from SEB-stimulated normal human blood run on mass cytometry (1.4, top), normal human bone marrow run on mass cytometry (5), normal human tonsil run on mass cytometry (2.5), SEB-stimulated normal human blood run on fluorescence flow cytometry (1.4, bottom), and a second analysis of normal human blood run on mass cytometry (6.1).

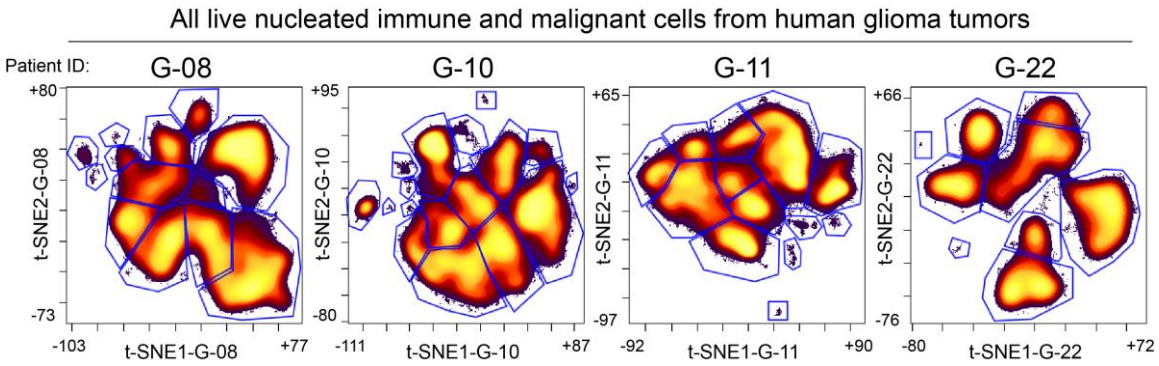

Figure 4-12. Unsupervised clustering and gating of 52 populations of malignant and immune cells in glioma. Live nucleated immune and malignant cells were gated from glioma tumors as described in Leelatian and Doxie et al., *Cytometry B* 2016(Leelatian, Doxie et al. 2016). Patient-specific t-SNE axes were created in separate viSNE analyses of each tumor (e.g. t-SNE1-G-08 for glioma tumor G-08). Shown here is density of cells on t-SNE1 vs. t-SNE2 from each tumor-specific viSNE analysis. Expert analysis of density was then used to identify 52 cell clusters from the 4 glioma tumors. These 52 populations were subsequently grouped by MEM in Fig. 4-13a using 9 proteins expressed on malignant cells.

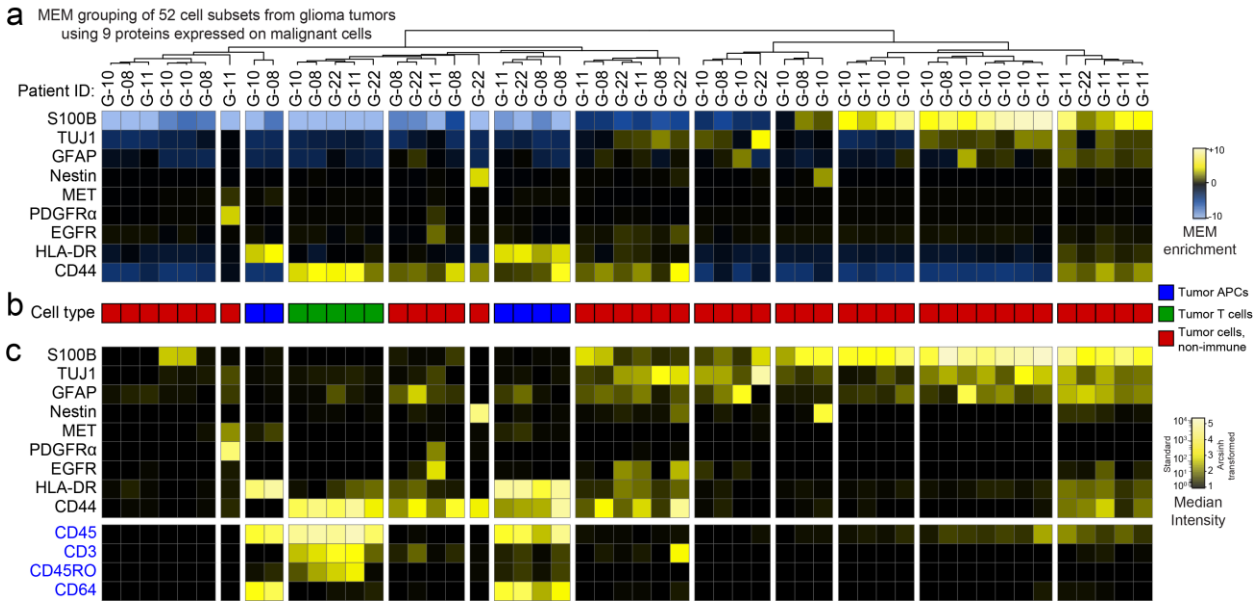

Figure 4-13. MEM correctly grouped immune and cancer cell populations from glioma tumors using nine proteins expressed on cancer cells. (A) A heatmap of MEM enrichment scores is shown for 52 populations of cells identified in tumors from 4 glioblastoma patients (G-08, G-10, G-11, G22) in an unsupervised manner using viSNE. MEM scores were then calculated based only on the nine measured proteins expected to be expressed on cancer cells (S100B, TJF1, GFAP, Nestin, MET, PGFRα, HLA-DR, and CD44). (B) Each population was annotated for a cell type based on review of the MEM label and classified as tumor infiltrating APCs (blue), tumor infiltrating T cells (green), or non-immune tumor cells (red). (C) A heatmap of median intensity values is shown for the 13 measured proteins from each of the 52 tumor cell populations. Expression of CD45, CD3, and CD64 was used to assess the respective identity of leukocytes, T cells, and antigen presenting cells.

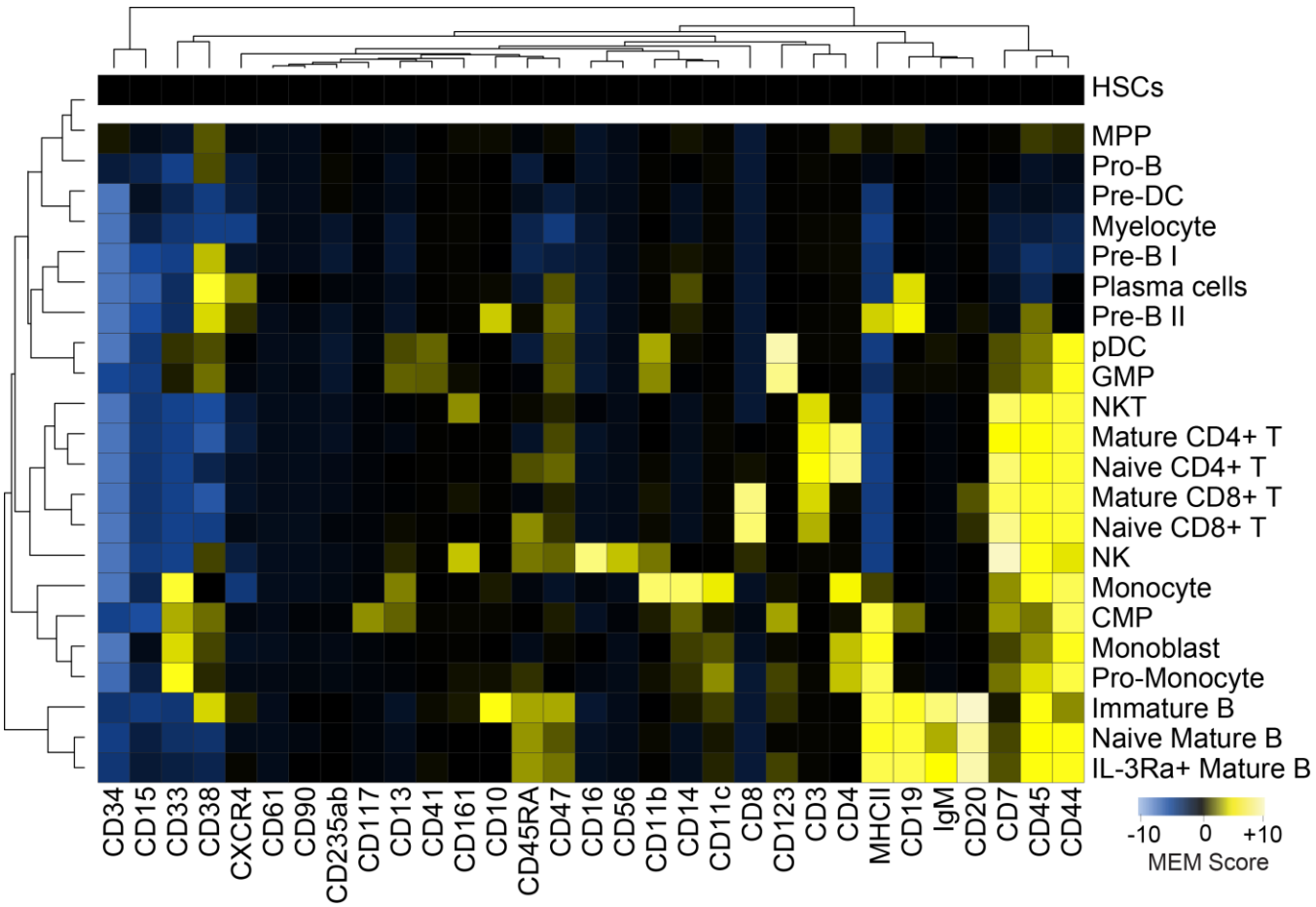

Figure 4-14. Using stem cells as reference population reveals high enrichment of maturation and differentiation markers on bone marrow cell subsets. Clustered heatmap of MEM scores from analysis using HSCs as the reference population for the bone marrow cell subsets identified by Bendall et al.(Bendall, Simonds et al. 2011) Comparison to stem cells as a common reference point provided intuitive MEM scores.

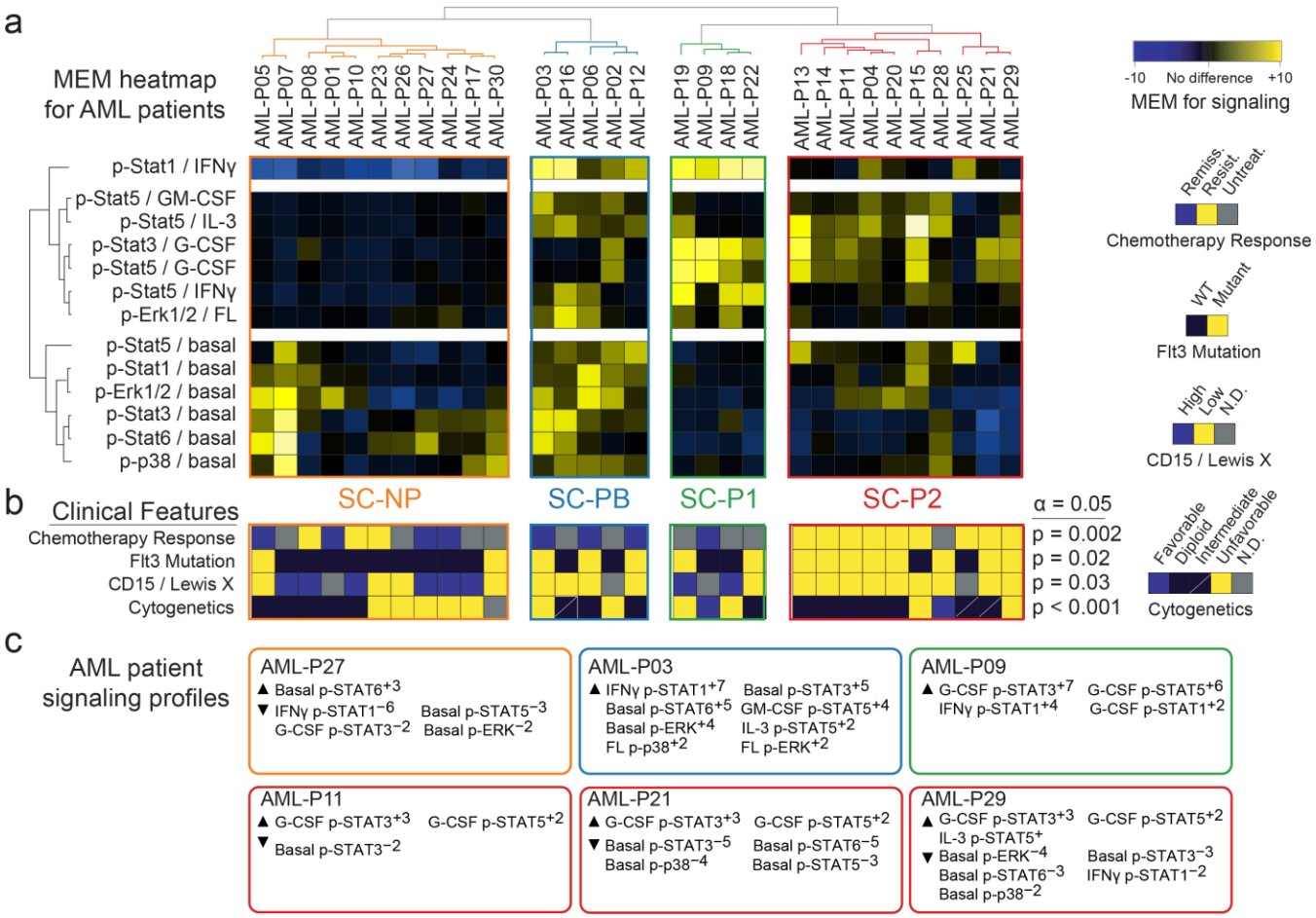

Figure 4-15. MEM labels reveal potentiated signaling in acute myeloid leukemia blasts from patients with poor chemotherapy responses. a) Previous analysis(Irish, Hovland et al. 2004) used median analysis of 2-color fluorescence cytometry and 3 staining panels to measure 6 phospho-protein responses to 6 stimulation conditions in leukemia blasts from 30 patients. The new heatmap here shows clustering of MEM scores for each patient compared to the other AML patients. Clustering on MEM enrichment scores identified the same four groups of patients as in the original analysis of median signaling response. b) The four patient cluster groups were defined based on basal and potential signaling and termed signaling-cluster non-potentiated (SC-NP), potentiated basal (SC-PB), potentiated-1 (SC-P1), or potentiated-2 (SC-P2), and SC-P2. SC-P2 was identified as enriched for high risk clinical features. Clinical features for each patient are shown(Irish, Hovland et al. 2004).c) Individual patient MEM scores are shown for 6 representative patients of the 30 AML patients.

Discussion

MEM provides a straightforward language for labeling cell types based on enriched features. Given widespread adoption and reporting, MEM labels could be used to identify and communicate cytotype across studies and platforms in a manner analogous to the way cluster of differentiation (CD) naming was used to communicate newly identified antigen targets of antibodies(1984). While it is useful in some cases to identify tissue-specific features of cell subsets, a more intuitive application of MEM compares populations against a common reference. As an example, HSCs were used as a common reference for all other cell subsets (Figure 4-11 and Figure 4-14).

This technique of using HSCs as a common reference was used to compare cell population similarity across experimental platforms, scientific groups, and tissue of origin. $CD4^+$ T cells were chosen due to the similarity in phenotype we might expect across tonsil, blood, and bone marrow, and it was striking that mass and fluorescence cytometry produced highly similar MEM labels (Table 4-1 and Figure 4-11). In contrast, SEB stimulated and unstimulated B cells measured by fluorescence cytometry had lower similarity scores (Figure 4-11). This arose because only 4 shared proteins were enriched on B cells measured by fluorescence cytometry, so the similarity scores were very sensitive to small changes in any one protein. This observation highlights the value of measuring more markers, especially in the setting of unusual or unexpected cell phenotypes. The more markers measured, the more chances for cells to be similar or different, and the more robust the MEM comparison. However, only 9 non-immune markers were needed to distinguish immune cells from cancer cells without including healthy immune cell markers to guide clustering (Figure 4-13). Going forward, MEM can be used to guide feature selection for both computational and experimental analysis.

Metrics like area under ROC curve (AUC), z-score, and K-S statistic quantify features of distributions that MEM does not capture, including difference or overlap, deviation, skew, and shape. AUC, for example, characterizes degree of non-overlap between two distributions and has recently been shown to be superior to K-S testing for characterizing “percent positive” in flow cytometry data(Kim, Donnenberg et al. 2016). Both K-S and AUC capture whether peaks are different. In some cases, difference metrics may be useful for highlighting features that are not enriched but which have different distribution shapes in the POP and REF. For example, bimodal vs. unimodal spreads with similar median and IQR values will score highly for K-S or AUC but have

near-zero MEM scores. This matches expert interpretation as we would say that neither population is enriched for the feature even though the distributions have differing shapes. MEM is set up as a scaled enrichment score so that it will be less sensitive to instrument variation and platform differences (as in Figure 4-11). MEM is also able to quantify and weight negative enrichment, an important aspect of population identity that is not always captured by other comparative metrics, such as z-score.

Other potential uses for MEM include clinical assessment of enriched features of patient response groups for the purpose of hypothesis generation and therapy development. For example, in an AML phospho-flow dataset(Irish, Hovland et al. 2004), MEM revealed potentiated STAT1 phosphorylation following IFNγ in AML cells from one patient group (Figure 4-15). IFNγ signaling via STAT1 activates expression of MHC and B7 family members including PD-L1, and potentiated signaling through this pathway might identify AML patients that could benefit from checkpoint inhibitor therapy(Loke and Allison 2003). MEM can also be used to monitor changes in tissues over time during treatment. Deviation from a stable MEM score for peripheral blood cell subsets would be expected in the case emerging malignant cells(Greenplate, Johnson et al. 2016), and lack of change towards a healthy set of MEM scores for blood or bone marrow cell subsets might indicate a lack of response to chemotherapy for a leukemia patient. Going forward, MEM is expected to assist in machine learning applications by providing quantitative text descriptions of cytotype that can be algorithmically parsed and used to classify newly identified cell subpopulations.

Methods

Data generation and acquisition

The tonsil and normal human bone marrow data sets (Table 4-1, indicated by *, and Figure 4-11) were generated by CyTOF analysis as described by Leelatian, et al(Leelatian, Diggins et al. 2015).

The normal human bone marrow data set from Bendall, et al(Bendall, Simonds et al. 2011) (Figure 4-6) was shared and downloaded from Cytobank as FCS files that included the cell population IDs defined by Bendall, et al(Bendall, Simonds et al. 2011).

The murine myeloid CyTOF dataset from Becher, et al(Becher, Schlitzer et al. 2014) (Figure 4-9) was shared and downloaded from Cytobank as FCS files that contained gated cell events and cluster IDs as designated by automated analysis in Becher et al, 2014(Becher, Schlitzer et al. 2014).

The phospho-flow AML data set(Irish, Hovland et al. 2004) (Figure 4-15) was shared on Cytobank and downloaded as FCS files.

The CyTOF datasets from analysis of human GBM, melanoma, and normal blood samples (Figure 4-13) was shared on Cytobank and downloaded as text files after viSNE analysis and gating. Data was previously analyzed and described in Leelatian et al, 2016(Leelatian, Doxie et al. 2016).

CyTOF data pre-processing and analysis

Data analysis was performed using the online analysis platform Cytobank(Kotecha, Krutzik et al. 2010) and the statistical programming environment R. Raw median intensity (MI) values were transformed to a hyperbolic arcsine scale. A cofactor of 15 was used for the PBMC dataset, and 5 was used for the normal human bone marrow data set and for the murine myeloid data set. Single, intact cells were gated based on cell length (30-60) and nucleic acid intercalator (iridium). Major PBMC subsets were gated based on CD45 expression (leukocytes) and on canonical lineage marker expression to identify major blood cell subsets.

FCS files were exported from Cytobank as FCS or tab-delimited text files that were parsed for expression intensity information in R(Team 2013). The R package flowCore was used to parse FCS files(Hahne, LeMeur et al. 2009). MEM was calculated using the arcsinh transformed MI values, as described above. Heatmaps were generated using the heatmap.2 function in the gplots package(Gregory R. Warnes 2015).

Fluorescence Phospho-Flow AML Data Analysis

Data were downloaded from Cytobank as FCS files and processed in R as described above. MFI values were transformed to a log normal scale. For each AML patient, a median value and an IQR value was calculated for each marker in the unstimulated condition and for the stimulated conditions. The unstimulated median values were subtracted from the stimulated median values, and likewise for the IQR values. MEM was then calculated by comparing each patient's subtracted median and IQR values to those of the other patients. This enabled a comparison of fold change signaling values rather than raw values.

Marker Enrichment Modeling (MEM)

MEM equation

The MEM equation is implemented as an R package. Currently, MEM uses medians as the magnitude value; however, depending on the data type, mean may be a more appropriate magnitude statistic and mean could be substituted for median in the equation. Similarly, other statistics, such as variance, might be substituted for IQR. The MEM equation was developed with the intention of capturing and quantifying population-specific feature enrichment in a simple equation that avoids over-fitting or unnecessary computation. The primary goal of this equation is to scale magnitude differences depending on distribution spread. While other distribution features such as skew or shape could be informative, incorporating only two pieces of information – magnitude and spread – into the equation captured enough information to be useful in quantifying both positive and negative population-specific feature enrichment.

MEM output and score scaling

The MEM R script outputs a heatmap of MEM values with a natural language summary of feature enrichment as the population (row) names. The + or - value provided along with the marker name is converted to a -10 to +10 scale and rounded to the nearest integer. As implemented here, the maximum of the scale was set using the highest absolute value MEM score observed across all markers and populations. All values in the matrix are divided by this maximum value and multiplied by 10 to achieve the -10 to +10 scaling. After scaling, the original sign value is reapplied to each MEM score. Scaling the output this way is intended to generate MEM values and labels that are intuitive to human readers and to facilitate comparison of feature enrichment across experiments, samples, batches, time points, and data types.

IQR Threshold

Because MEM uses a ratio of IQR values, near zero values in the denominator, IQR_{POP}, will greatly increase MEM scores. For each measurement type, it is important to identify a minimum significant IQR value so that small IQR values below the platform's ability to distinguish signal from noise do not inappropriately increase MEM scores. To automatically determine a minimum threshold for IQR_{POP}, the algorithm here calculated

the average of the IQR values that were associated with the lowest quartile of population and reference medians. For the mass and fluorescence cytometry datasets used, the automatically calculated IQR threshold was on average 0.5 ± X and so the IQR threshold for all studies here was set to 0.5. The default IQR threshold in the algorithm is also set to 0.5. To have the IQR threshold re-calculated, investigators should specify the "auto" option for the IQR.thresh argument in the MEM function. It is recommended that investigators applying MEM to datasets from different instruments or who are testing MEM for the first time determine whether a change in the IQR threshold is needed.

Reference population selection

MEM score are contextual in that a population's MEM score depends on the reference population(s) to which it is compared. Selection of a reference population should be made deliberately depending on the biological question being addressed. When populations in a MEM analysis arise from different experimental sources, it may be necessary in some cases to normalize measurements prior to MEM analysis to avoid artifacts from experimental variation.

PBMC processing and mass cytometry

Peripheral blood mononuclear cells (PBMC) were isolated and cryopreserved as in Greenplate, et al(Greenplate, Johnson et al. 2016). PBMC were stained with metal conjugated antibodies and prepared for the mass cytometry as previously described(Greenplate, Johnson et al. 2016). The following antibodies were used in the staining panel: CD19-142, CCR5-144, CD4-145, CD64-146, CD20-147, CCR4-149, CD43-150, CD14-151, TCRγδ-152, CD45RA-153, CD45-154, CXCR3-156, CD33-158, CCR7-159, CD28-169, CD29-162, CD45RO-164, CD16-165, CD44-166, CD27-167, CD8-168, CD25-169, CD3-170, CD57-172, PD-L1-175, and CD56-176 (Fluidigm Sciences). In addition, the following purified antibodies from Biolegend were labeled using MaxPar DN3 kits (Fludigm Sciences), stored at 4°C in antibody stabilization buffer (Candor Bioscience GmbH) and used in the same panel: ICOS-141, TIM-143, CD38-148, CD32-161, HLA-DR-163, CXCR5-171, and PD-1-174.

Cell subpopulation MEM Score Similarity Calculations

Comparison of $CD4^+$ T cells to B cells

In order to assess the robustness of MEM across tissue sample types, donors, experimental runs, and flow cytometry platforms (fluorescence and mass cytometry), MEM scores were calculated for cell subsets from 7 different experiments that included 3 healthy human bone marrow samples(Bendall, Simonds et al. 2011, Amir el, Davis et al. 2013, Ferrell, Diggins et al. 2016), 9 healthy human PBMC samples(Leelatian, Diggins et al. 2015, Nicholas, Greenplate et al. 2016), and 6 healthy human tonsil samples(Polikowsky, Wogsland et al. 2015). MEM scores were calculated for each population using as the reference population a combination of hematopoietic stem cells gated as $CD34^{+}$ $CD38^{lo/-}$ from two studies of healthy human bone marrow(Bendall, Simonds et al. 2011, Ferrell, Diggins et al. 2016). Population similarity was calculated using root mean squared distance (RMSD) calculated on all population MEM scores in a pairwise fashion. MEM scores were calculated using all markers in common between each dataset and the HSC reference (Table 4-5).

Table 4-5. Full antibody panels for immune cell datasets in Fig. 4

Dataset	Panel
Nicholas KJ et al.(Nicholas, Greenplate et al. 2016)	HLADR, CD8, CD25, CD38, CD4, CD3, CD62L, CD69
Nicholas KJ et al.(Nicholas, Greenplate et al. 2016)	HLADR, CD19, CD27, CD38, CD86, CD20
Greenplate AR[§]	ICOS, CD19, TIM3, CCR5, CD4, CD64, CD20, CD38, CCR4, CD43, CD14, TCRγδ, CD45RA, CD45, CXCR3, CD33, CCR7, CD28, CD32, CD69, HLADR, CD45RO, CD16, CD44, CD27, CD8, CD2, CD3, CXCR5, CD57, PD1, PDL1, CD56
Ferrell et al.(Ferrell, Diggins et al. 2016)	CD235a, CD19, CD117, CD11b, CD4, CD64, CD7 , CD34, CD61, CD123, CD13, CD62L, CD45, CD183, CD33, CD11c, CD14, CD15, CD16, CD24, CD38, CD25, CD3, CD185, HLA-DR, CD184, CD56
Leelatian et al.(Leelatian, Diggins et al. 2015)	CD19, CD117, CD11b, CD4, CD8a, CD20, CD34, CD61, CD123, CD45RA, CD45, CD10, CD33, CD11c, CD14, CD69, CD15, CD16, CD44, CD38, CD25, CD3, IgM, HLADR, CD56
Bendall et al.(Bendall, Simonds et al. 2011)	CD45, CD45RA, CD235ab, CD19, CD11b, CD4, CD8, CD34, CD161, CD20, CD41, CD11c, CD123, IgM, CD10, CD33, CD14, CD38, CD15, CD16, CD44, CD7, CD13, CD56, CD61, CD117, CD47, HLADR, CD90, CXCR4, CD3
Polikowsky et al.(Polikowsky, Wogsland et al. 2015)	CD19, CD5, IgG, CD4, IgD, CD20, CD16, Igλ, CD45, CD27, CD86, CD33, CD22, Igκ, CD79B, CD40, CD44, CD38, CD8, CD3, IgM, HLADR, SHP1, CD56
Amir et al.(Amir el, Davis et al. 2013)	CD45, CD3, CD45RA, CD19, CD11b, CD4, CD8, CD34, CD20, Ki67, CD33, CD123, IκBα, CD38, CD90
HSC REF**:** Amir et al.(Amir el, Davis et al. 2013) & Ferrell et al.(Ferrell, Diggins et al. 2016)	CD45, CD45RA, CD235ab, CD19, CD11b, CD4, CD8, CD34, CD161, CD20, CD41, CD11c, CD123, IgM, CD10, CD33, CD14, CD38, CD15, CD16, CD44, CD7, CD13, CD56, CD61, CD117, CD47, HLADR, CD90, CXCR4

Note: MEM comparisons were made using all markers in common between each dataset and the HSCs (combined from Amir et al. and Ferrell et al. datasets). Pairwise RMSD comparisons of MEM scores were made using all markers in common between the pairs of datasets. [§]Unpublished data from AR Greenplate, prepared as in Leelatian et al.(Leelatian, Diggins et al. 2015)

RMSD was calculated here as the square root of the average in squared distance between all MEM values in common for each pair of populations (Table 4-4) and then converted into percent maximum possible RMSD. Given the -10 to 10 MEM scale, an RMSD of 20 was the maximum possible difference and corresponded to 0% similarity, whereas an RMSD of 0 between MEM labels indicated 100% similarity. This approach emphasized differences in marker expression when comparing populations. Calculated statistics for $CD4^+$ T cell comparisons included average MEM value +/- standard deviation and p-value calculated using an unpaired Student's t-test.

Human Glioma and Normal Immune Cell MEM Analysis

Glioblastoma data (G-08, G-10, G-11, and G-22) were collected following a published protocol(Leelatian, Doxie et al. 2016). Cells were stained with isotope-tagged antibodies to detect surface and intracellular targets following established protocols(Leelatian, Diggins et al. 2015, Leelatian, Doxie et al. 2016). MEM analysis of glioblastoma patient samples were performed with 9 markers (S100B, TUJ1, GFAP, Nestin, MET, PDGFRα, EGFR, HLA-DR, and CD44), using arcsinh transformation of original median intensity values with a cofactor of 5. Each cell subset was the POP, and the remaining cell subsets were the REF in the analysis.

Z-score and K-S statistic calculations

Z-score was calculated between POP and REF as (MEANpop-MEANref)/STDEVref for each marker.

The K-S statistic was calculated comparing the distribution for each marker on POP and REF using the function ks.test() in R.

F-measure Analysis

PBMC populations were defined by expert human gating on canonical markers. For f-measure analysis (Figure 4-2c and Figure 4-5a), the 25 measured markers from the CyTOF analysis of healthy PBMC were sorted based on absolute MEM scores, median values, median difference (shown in Figure 4-5a), or randomly across all PBMC populations and the 25 measured proteins. The 5x25 matrix was converted into an ordered vector (length 25X5) and then sorted by absolute value. The first occurrence of each marker in the list was kept and

subsequent occurrences of that marker in the list (i.e. that marker's scores on other populations) were discarded. The order of markers excluded by MEM, median, median difference, z-score, and K-S statistic are shown in Table 4-3. Markers were then sequentially, cumulatively excluded from k-means clustering of cells from high to low absolute for each statistic or score. F-measure was calculated as:

Sensitivity = True Positives/ (True Positives + False Negatives)

Specificity = True Negatives/ (True Negatives + False Positives)

F-measure = 2*(sensitivity*specificity)/ (sensitivity + specificity)

An F-measure was calculated for each round of clustering, where truth was the cell cluster ID resulting from clustering on all 25 markers. The moving average of f-measure with an interval of 3 was calculated in Microsoft Excel. The f-measures for random marker exclusion are the average at each point of 15 different rounds of random marker exclusion from clustering.

Acknowledgements

This study was supported by R25 CA136440-04 (K.E.D.), R00 CA143231-03 (J.M.I.), the Vanderbilt-Ingram Cancer Center (VICC, P30 CA68485), and VICC Young Ambassadors and VICC Hematology Helping Hands awards. Thanks to Mikael Roussel for helpful discussions of myeloid cell identity markers, to Deon Doxie for helpful discussions of MEM analysis of tumor and immune cell subsets, and to Lola Chambless and Rebecca Ihrie for use of glioma tumor data generated by N.L.

Conflicts of Interest: JMI has a financial interest as co-founder and board member in Cytobank Inc., a software company for single cell data analysis. No other conflicts.

Author contribution

All authors designed experiments, discussed data visualization and contributed intellectually to the manuscript, and approved the final manuscript. J.M.I. and K.E.D. performed computational analyses, developed analytical tools and protocols, conceived and designed the study, and wrote the manuscript. A.R.G. contributed

to Fig. 4 and Fig. 5 and assisted with manuscript revisions. N.L. contributed to Fig. 5 and manuscript revisions. C.E.W. contributed to R code implementation and manuscript revisions.

CHAPTER 5

CONCLUSION AND FUTURE DIRECTIONS

A single-cell level understanding of cancer cells and the microenvironment is essential for identifying, quantitatively describing, and tracking biologically and clinically relevant populations of cells. Advances in computational analysis tools and workflows enable single-cell level characterization of tissues in order to better understand human health and disease.

This dissertation presents and demonstrates a modular analysis workflow that combines supervised and unsupervised computational analysis for automated population discovery and a comprehensive view of immune and AML cell phenotypes. The initial biaxial gating and dimensionality reduction allows the analyst to assess data quality and provides a complete overview of the data. However, grouping cells into subpopulations in the last steps of the workflow is completely automated, thereby providing a robust and repeatable analysis approach that can be scaled across large datasets and studies. Going forward, it will be beneficial to develop standardized, modular workflows that leverage the utilities of multiple computational tools in order to extract as much meaningful data as possible from these types of multi-dimensional data.

The tools that have been developed for population identification, such as dimensionality reduction, can also also be used as the basis for novel quantitative analyses, as demonstrated in Chapter 3. This study highlighted the utility of mass cytometry and computational analysis tools for tracking changes in AML blast cell populations over time and after treatment. The use of viSNE analysis was shown to be an efficient means of tracking AML cell populations after treatment in order to determine which cells persisted after therapy. This analysis method revealed that these persister cells became less phenotypically stem-like following treatment. This finding was unexpected given that cancer stem cells have been shown to be responsible for relapse in AML and would therefore be expected to express surface proteins similarly to normal stem cells.

Chapter 4 presents MEM, an automated method for quantifying feature enrichment, and demonstrates how this tool is able to identify proteins that are important to cell population identity in multiple datasets from analysis of immune cells and cancers. Markers known to be positively and negatively enriched on specific populations of human and murine immune cells from multiple tissue types were highlighted by MEM analysis.

MEM labels also were successfully used to group CD4+ T cells and B cells from different tissues, experiments, and platforms. The possibility of using MEM labels to group similar populations of cells holds great potential for future applications in machine learning approaches to register cells between studies, particularly in cases where biologically or clinically relevant populations of cells have been identified on multiple platforms where direct comparison of the data is not possible due to highly divergent ranges and scales. MEM also successfully distinguished populations of human immune cells from glioma cells even when common immune cell markers were left out of the analysis.

Going forward, these methods can be adapted and applied for automated cell population discovery, tracking, and quantitative comparison. The modular nature of the computational analysis workflow presented here lends itself to adaptations for various experimental and clinical applications. For example, alternative dimensionality reduction methods like ISOMAP or LLE, described in Chapter 2, could be substituted for viSNE analysis for data expected to have show continuous, progressive phenotypes in contrast to the discrete populations found in samples like human blood. Alternative clustering or automated gating methods could also be substituted for SPADE. This overall workflow, which broadly entails quality control, comprehensive data visualization for major group identification, and automated subpopulation identification, can also be extrapolated to other data types, such gene expression and imaging data. Another key potential area of advancement for this workflow is increased automation for each step. Highly automated workflows will expedite analysis of large datasets and produce more robust and repeatable results.

The method outlined in Chapter 3 for quantifying cellular phenotypic distance in a reduced-dimensional space establishes the conceptual framework for quantifying sample-level cellular heterogeneity. The concept presented here highlights the need for multi-variate, quantitative metrics of phenotypic heterogeneity. Tools like viSNE and SPADE provide informative qualitative representations of the data. However, quantitative approaches are needed to comprehensively compare sample-level heterogeneity and correlate this heterogeneity with clinical outcomes and biological mechanisms. When quantified, cellular heterogeneity has the potential to serve as a biomarker of disease prognosis, therapy response, and other clinical parameters.

Advancements in cancer therapy rely on the development of novel methods for quantifying the features specific to clinically relevant populations of cells, for tracking those populations of cells over the course of therapy

and recovery, and comparing populations of cells from different patients, studies, and tissue sources. MEM analysis provides scaled enrichment scores in the form of labels that are comparable between platforms and robust against instrument variation. These labels can be used as quantitative descriptors for cell populations and compared across studies to automatically determine population identities. Going forward, a database could be created to store the MEM labels derived from established cell types against which newly identified populations can be compared. A machine learning algorithm could also be developed to improve cell classification as additional MEM labels were added to the database. Given a population of cells and set of measured features, the tool would calculate a MEM score for that population using either a user-defined population or a standardized cell population as reference. The resulting MEM scores could then be compared to the MEM labels of previously defined cell populations in the database, and probability or similarity scores would be returned estimating the most likely biological identity of the new population of cells.

The methods described in this dissertation also have significant potential for clinical applications. As high dimensional analysis of clinical samples becomes routine, workflows such as that described in Chapter 2 will be needed to rapidly and automatically group cells into subpopulations and provide an interpretable visual output for clinical assessment. These types of tools may also be useful for identifying cellular features that correspond to clinical parameters, as demonstrated by the use of viSNE in Chapter 3 of this dissertation for quantifying a shift in AML blast phenotype that is seen in patients with negative outcomes. Tools that quantify cellular heterogeneity and feature enrichment therefore have the potential to provide novel biomarkers for disease prognosis. MEM analysis may also be useful for identifying features specifically enriched on clinically relevant populations of cells, specifically those resistant to therapy, and point to potential therapeutic targets against the resistant cell populations.

As technological advancements lead to increasingly large datasets, novel computational tool development will be needed to extract meaningful and useful information from these data in order to continue improving disease treatment and gain a more complete understanding of human biology.

APPENDIX A

USE OF VISNE ANALYSIS TO COMPARE METABOLISM OF CELL POPULATIONS

Preface

While the majority of my dissertation research focused on the analysis and interpretation of flow cytometry data, many other single-cell experimental platforms are available that provide valuable single-cell level information about biological samples. The ability to apply computational analysis tools and workflows to data generated across platforms will ultimately enable a more direct comparison of cells discovered using different methods. As a use case, I collaborated with Dr. Amy Shah, then a graduate student in Melissa Skala's lab at Vanderbilt University, to apply computational analysis tools that were originally developed for flow cytometry data to single-cell imaging data. My goal was to use these tools to gain a comprehensive understanding of single-cell metabolic phenotypes of cells from a mouse model of head and neck cancer. To accomplish this, I linearly scaled the imaging parameters to a common 0-100 scale and then applied viSNE analysis to the scaled data in Cytobank (www.cytobank.org) to visualize the cells in a two-dimensional map according to their metabolic parameter similarities. My analysis ultimately confirmed that there was a greater degree of metabolic heterogeneity in the samples treated with chemotherapeutic drugs. Adapting tools like viSNE for use across experimental platforms will enable more efficient tracking of cell subsets identified from different tissues, studies, and platforms, thereby expanding

The work presented here was originally published in *Neoplasia* 2015 (open access) (Shah, Diggins et al. 2015). http://www.neoplasia.com/article/S1476-5586(15)00143-8/abstract

Introduction

Cellular metabolism is altered in cancer cells, resulting in an abnormal ratio of cellular metabolites that can be experimentally measured. Two metabolic cofactors frequently used to determine the metabolic state of the cell, NAD(P)H and FAD, autofluorescence at measurable wavelengths and can therefore be imaged using high resolution autofluorescence imaging. The ratio of NAD(P)H and FAD fluorescence intensity, referred to as the optical redox ratio, is an established measure of cellular metabolism. In order to assess how two common

cancer drugs, cetuximab and cisplatin, affected cancer cell metabolism in a mouse model of head and neck squamous cell carcinoma, mice with FaDu tumors were treated with the compounds and the tumors were subsequently imaged to quantify the per-cell optical redox ratio. To visualize the heterogeneity in cellular metabolism, viSNE analysis was used to organize cells in a reduced-dimensional space according to imaging parameters. From these analyses, a greater degree of cellular metabolic heterogeneity was found in tumors treated with cetuximab and cisplatin compared to an untreated control.

Methods

Imaging analysis of mouse tumor was performed as described in Shah, et al (Shah, Diggins et al. 2015). The seven optical metabolic imaging parameters that were measured included NAD(P)H α_1, NAD(P)H τ_1, NAD(P)H τ_2, FAD α_2, FAD τ_1, and FAD τ_2. In order to compare the parameters, they were each transformed to a common scale from 0 to 100, with the highest value for each parameter being 100. The data were converted from Excel matrices to FCS files that could be uploaded to and analyzed using Cytobank (www.cytobank.org) for viSNE analysis. All seven of the scale-transformed parameters were used in the viSNE analysis.

Results

viSNE analysis of all seven measured optical metabolic imaging parameters revealed a population of cells from the cetuximab and treatment groups that was distinct from the cells of the control group (Figure A-1A). Visualization of heat intensity of the imaging parameters on the viSNE ap indicated that NAD(P)H τ_1 and FAD τ_1 contributed the most variation to the mapping of the cells (Figure A-1B and C). The increased cellular heterogeneity seen in the treatment group by viSNE analysis was also seen in analysis of individual imaging parameters (Shah, Diggins et al. 2015).

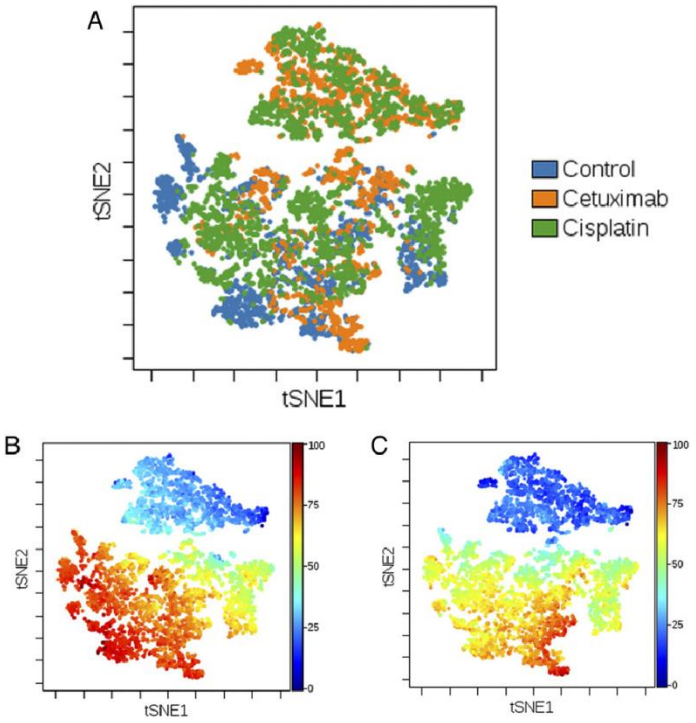

Figure A-1. Single-cell analysis using the dimensionality-reduction technique viSNE reduced seven optical metabolic imaging parameters to two dimensions for visualization of heterogeneity across individual cells. To account for different scales between parameters, common linear transformation was applied within each parameter across all treatment groups so the transformed values ranged from 0 to 100. A) viSNE analysis shows a distinct population of cells for the control group. The cetuximab and cisplatin treatment groups overlap with the control group and also comprise a separate population of cells (B,C). Heat intensity indicates short fluorescence lifetime components ($\tau 1$) for (B) NAD(P)H and (C) FAD.

Conclusions

viSNE analysis was successfully applied to single-cell fluorescence imaging data to reveal the metabolic cellular heterogeneity that arose from cisplatin and cetuximab treatment of mice with FaDu tumors. Dimensionality reduction tools like viSNE can be used with single-cell data from non-flow cytometry platforms to inform on cellular heterogeneity and to identify distinct populations of cells.

BIBLIOGRAPHY

(1984). "Nomenclature for clusters of differentiation (CD) of antigens defined on human leukocyte populations. IUIS-WHO Nomenclature Subcommittee." Bull World Health Organ **62**(5): 809-815.

(2013). Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) Research Data (1973-2011), National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch, released April 2014, based on the November 2013 submission.

Abdelrahman, A. I., S. Dai, S. C. Thickett, O. Ornatsky, D. Bandura, V. Baranov and M. A. Winnik (2009). "Lanthanide-containing polymer microspheres by multiple-stage dispersion polymerization for highly multiplexed bioassays." J Am Chem Soc **131**(42): 15276-15283.

Abdelrahman, A. I., O. Ornatsky, D. Bandura, V. Baranov, R. Kinach, S. Dai, S. C. Thickett, S. Tanner and M. A. Winnik (2010). "Metal-Containing Polystyrene Beads as Standards for Mass Cytometry." J Anal At Spectrom **25**(3): 260-268.

Aghaeepour, N., G. Finak, C. A. P. C. Flow, D. Consortium, H. Hoos, T. R. Mosmann, R. Brinkman, R. Gottardo and R. H. Scheuermann (2013). "Critical assessment of automated flow cytometry data analysis techniques." Nat Methods **10**(3): 228-238.

Amir el, A. D., K. L. Davis, M. D. Tadmor, E. F. Simonds, J. H. Levine, S. C. Bendall, D. K. Shenfeld, S. Krishnaswamy, G. P. Nolan and D. Pe'er (2013). "viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia." Nat Biotechnol **31**(6): 545-552.

Arellano, M., S. Pakkala, A. Langston, M. Tighiouart, L. Pan, Z. Chen, L. T. Heffner, S. Lonial, E. Winton and H. J. Khoury (2012). "Early clearance of peripheral blood blasts predicts response to induction chemotherapy in acute myeloid leukemia." Cancer **118**(21): 5278-5282.

Baer, M. R., C. C. Stewart, R. K. Dodge, G. Leget, N. Sule, K. Mrozek, C. A. Schiffer, B. L. Powell, J. E. Kolitz, J. O. Moore, R. M. Stone, F. R. Davey, A. J. Carroll, R. A. Larson and C. D. Bloomfield (2001). "High frequency of immunophenotype changes in acute myeloid leukemia at relapse: implications for residual disease detection (Cancer and Leukemia Group B Study 8361)." Blood **97**(11): 3574-3580.

Bagwell, C. B., J. L. Hudson and G. L. Irvin, 3rd (1979). "Nonparametric flow cytometry analysis." J Histochem Cytochem **27**(1): 293-296.

Balkay, L. "FCS data reader (http://www.mathworks.com/matlabcentral/fileexchange/9608-fcs-data-reader)." MATLAB Central File Exchange **Retrieved June 2014**.

Bandura, D. R., V. I. Baranov, O. I. Ornatsky, A. Antonov, R. Kinach, X. Lou, S. Pavlov, S. Vorobiev, J. E. Dick and S. D. Tanner (2009). "Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry." Anal Chem **81**(16): 6813-6822.

Basit, A., J. Reutershan, M. A. Morris, M. Solga, C. E. Rose, Jr. and K. Ley (2006). "ICAM-1 and LFA-1 play critical roles in LPS-induced neutrophil recruitment into the alveolar space." Am J Physiol Lung Cell Mol Physiol **291**(2): L200-207.

Becher, B., A. Schlitzer, J. Chen, F. Mair, H. R. Sumatoh, K. W. W. Teng, D. Low, C. Ruedl, P. Riccardi-Castagnoli, M. Poidinger, M. Greter, F. Ginhoux and E. W. Newell (2014). High-dimensional analysis of the murine myeloid cell system. Nature Immunology. **15:** 1181-1189.

Bendall, S. C., K. L. Davis, A. D. Amir el, M. D. Tadmor, E. F. Simonds, T. J. Chen, D. K. Shenfeld, G. P. Nolan and D. Pe'er (2014). "Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development." Cell **157**(3): 714-725.

Bendall, S. C., E. F. Simonds, P. Qiu, A. D. Amir el, P. O. Krutzik, R. Finck, R. V. Bruggner, R. Melamed, A. Trejo, O. I. Ornatsky, R. S. Balderas, S. K. Plevritis, K. Sachs, D. Pe'er, S. D. Tanner and G. P. Nolan (2011). "Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum." Science **332**(6030): 687-696.

Bengtsson, H. (2015). "R.utils: Various Programming Utilities." http://cran.r-project.org/web/packages/R.utils/R.utils.pdf.

Berraondo, P., L. Minute, D. Ajona, L. Corrales, I. Melero and R. Pio (2016). "Innate immune mediators in cancer: between defense and resistance." Immunol Rev **274**(1): 290-306.

Bruggner, R. V., B. Bodenmiller, D. L. Dill, R. J. Tibshirani and G. P. Nolan (2014). "Automated identification of stratifying signatures in cellular subpopulations." Proc Natl Acad Sci U S A **111**(26): E2770-2777.

Chen, X., M. Hasan, V. Libri, A. Urrutia, B. Beitz, V. Rouilly, D. Duffy, E. Patin, B. Chalmond, L. Rogge, L. Quintana-Murci, M. L. Albert, B. Schwikowski and C. Milieu Interieur (2015). "Automated flow cytometric analysis across large numbers of samples and cell types." Clin Immunol.

Civin, C. I., L. C. Strauss, C. Brovall, M. J. Fackler, J. F. Schwartz and J. H. Shaper (1984). "Antigenic analysis of hematopoiesis. III. A hematopoietic progenitor cell surface antigen defined by a monoclonal antibody raised against KG-1a cells." J Immunol **133**(1): 157-165.

Cox, C., J. E. Reeder, R. D. Robinson, S. B. Suppes and L. L. Wheeless (1988). "Comparison of frequency distributions in flow cytometry." Cytometry **9**(4): 291-298.

Dalerba, P., S. J. Dylla, I. K. Park, R. Liu, X. Wang, R. W. Cho, T. Hoey, A. Gurney, E. H. Huang, D. M. Simeone, A. A. Shelton, G. Parmiani, C. Castelli and M. F. Clarke (2007). "Phenotypic characterization of human colorectal cancer stem cells." Proc Natl Acad Sci U S A **104**(24): 10158-10163.

Diggins, K. E., P. B. Ferrell, Jr. and J. M. Irish (2015). "Methods for discovery and characterization of cell subsets in high dimensional mass cytometry data." Methods **82**: 55-63.

Ding, L., T. J. Ley, D. E. Larson, C. A. Miller, D. C. Koboldt, J. S. Welch, J. K. Ritchey, M. A. Young, T. Lamprecht, M. D. McLellan, J. F. McMichael, J. W. Wallis, C. Lu, D. Shen, C. C. Harris, D. J. Dooling, R. S. Fulton, L. L. Fulton, K. Chen, H. Schmidt, J. Kalicki-Veizer, V. J. Magrini, L. Cook, S. D. McGrath, T. L. Vickery, M. C. Wendl, S. Heath, M. A. Watson, D. C. Link, M. H. Tomasson, W. D. Shannon, J. E. Payton, S. Kulkarni, P. Westervelt, M. J. Walter, T. A. Graubert, E. R. Mardis, R. K. Wilson and J. F. DiPersio (2012). "Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing." Nature **481**(7382): 506-510.

Doulatov, S., F. Notta, E. Laurenti and J. E. Dick (2012). "Hematopoiesis: a human perspective." Cell Stem Cell **10**(2): 120-136.

Elliott, M. A., M. R. Litzow, L. L. Letendre, R. C. Wolf, C. A. Hanson, A. Tefferi and M. S. Tallman (2007). "Early peripheral blood blast clearance during induction chemotherapy for acute myeloid leukemia predicts superior relapse-free survival." Blood **110**(13): 4172-4174.

Ellis B, G. R., Hahne F, Meur NL and Sarkar D "flowViz: Visualization for flow cytometry." R package version 1.30.0.

Ellis B, H. P., Hahne F, Meur NL, Gopalakrishnan N, Spidlen J "flowCore: flowCore: Basic structures for flow cytometry data." R package version 1.32.2.

Eppert, K., K. Takenaka, E. R. Lechman, L. Waldron, B. Nilsson, P. van Galen, K. H. Metzeler, A. Poeppl, V. Ling, J. Beyene, A. J. Canty, J. S. Danska, S. K. Bohlander, C. Buske, M. D. Minden, T. R. Golub, I. Jurisica, B. L. Ebert and J. E. Dick (2011). "Stem cell gene expression programs influence clinical outcome in human leukemia." Nat Med **17**(9): 1086-1093.

Ferrell, P. B., Jr., K. E. Diggins, H. G. Polikowsky, S. R. Mohan, A. C. Seegmiller and J. M. Irish (2016). "High-Dimensional Analysis of Acute Myeloid Leukemia Reveals Phenotypic Changes in Persistent Cells during Induction Therapy." PLoS One **11**(4): e0153207.

Fienberg, H. G., E. F. Simonds, W. J. Fantl, G. P. Nolan and B. Bodenmiller (2012). "A platinum-based covalent viability reagent for single-cell mass cytometry." Cytometry A **81**(6): 467-475.

Finak, G., J. Frelinger, W. Jiang, E. W. Newell, J. Ramey, M. M. Davis, S. A. Kalams, S. C. De Rosa and R. Gottardo (2014). "OpenCyto: An Open Source Infrastructure for Scalable, Robust, Reproducible, and Automated, End-to-End Flow Cytometry Data Analysis." PLoS Comput Biol **10**(8): e1003806.

Finck, R., E. F. Simonds, A. Jager, S. Krishnaswamy, K. Sachs, W. Fantl, D. Pe'er, G. P. Nolan and S. C. Bendall (2013). "Normalization of mass cytometry data with bead standards." Cytometry A **83**(5): 483-494.

Furze, R. C. and S. M. Rankin (2008). "Neutrophil mobilization and clearance in the bone marrow." Immunology **125**(3): 281-288.

Geoffrey Hinton, S. R. (2002). "Stochastic neighbor embedding." Advances in neural information processing systems.

Gianfaldoni, G., F. Mannelli, M. Baccini, E. Antonioli, F. Leoni and A. Bosi (2006). "Clearance of leukaemic blasts from peripheral blood during standard induction treatment predicts the bone marrow response in acute myeloid leukaemia: a pilot study." Br J Haematol **134**(1): 54-57.

Giesen, C., H. A. Wang, D. Schapiro, N. Zivanovic, A. Jacobs, B. Hattendorf, P. J. Schuffler, D. Grolimund, J. M. Buhmann, S. Brandt, Z. Varga, P. J. Wild, D. Gunther and B. Bodenmiller (2014). "Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry." Nat Methods.

Goardon, N., E. Marchi, A. Atzberger, L. Quek, A. Schuh, S. Soneji, P. Woll, A. Mead, K. A. Alford, R. Rout, S. Chaudhury, A. Gilkes, S. Knapper, K. Beldjord, S. Begum, S. Rose, N. Geddes, M. Griffiths, G. Standen, A. Sternberg, J. Cavenagh, H. Hunter, D. Bowen, S. Killick, L. Robinson, A. Price, E. Macintyre, P. Virgo, A. Burnett, C. Craddock, T. Enver, S. E. Jacobsen, C. Porcher and P. Vyas (2011). "Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia." Cancer Cell **19**(1): 138-152.

Greenplate, A. R., D. B. Johnson, P. B. Ferrell and J. M. Irish (2016). "Systems immune monitoring in cancer therapy." Eur J Cancer **61**: 77-84.

Greenplate, A. R., D. B. Johnson, M. Roussel, M. R. Savona, J. A. Sosman, I. Puzanov, P. B. Ferrell and J. M. Irish (2016). "Myelodysplastic Syndrome Revealed by Systems Immunology in a Melanoma Patient Undergoing Anti-PD-1 Therapy." Cancer Immunol Res.

Gregory R. Warnes, B. B., Lodewijk Bonebakker, Robert Gentleman, Wolfgang Huber Andy Liaw, Thomas Lumley, Martin Maechler, Arni Magnusson, Steffen Moeller, Marc Schwartz, Bill Venables (2015). "gplots: Various R Programming Tools for Plotting Data."

Hadrup, S., M. Donia and P. Thor Straten (2013). "Effector CD4 and CD8 T cells and their role in the tumor microenvironment." Cancer Microenviron **6**(2): 123-133.

Hahne, F., N. LeMeur, R. R. Brinkman, B. Ellis, P. Haaland, D. Sarkar, J. Spidlen, E. Strain and R. Gentleman (2009). "flowCore: a Bioconductor package for high throughput flow cytometry." BMC Bioinformatics **10**: 106.

Hanahan, D. and R. A. Weinberg (2011). "Hallmarks of cancer: the next generation." Cell **144**(5): 646-674.

Herzenberg, L. A., J. Tung, W. A. Moore, L. A. Herzenberg and D. R. Parks (2006). "Interpreting flow cytometry data: a guide for the perplexed." Nat Immunol **7**(7): 681-685.

Inokuma, M. S., V. C. Maino and C. B. Bagwell (2013). "Probability state modeling of memory CD8(+) T-cell differentiation." J Immunol Methods **397**(1-2): 8-17.

Irish, J., R. Hovland, P. Krutzik, O. Perez, O. Bruserud, B. Gjertsen and G. Nolan (2004). "Single cell profiling of potentiated phospho-protein networks in cancer cells." Cell **118**(2): 217-228.

Irish, J., N. Kotecha and G. Nolan (2006). "Innovation - Mapping normal and cancer cell signalling networks: towards single-cell proteomics." Nature Reviews Cancer **6**(2): 146-155.

Irish, J. M. (2014). "Beyond the age of cellular discovery." Nat Immunol **15**(12): 1095-1097.

Irish, J. M. and D. B. Doxie (2014). "High-dimensional single-cell cancer biology." Curr Top Microbiol Immunol **377**: 1-21.

Irish, J. M., R. Hovland, P. O. Krutzik, O. D. Perez, O. Bruserud, B. T. Gjertsen and G. P. Nolan (2004). "Single cell profiling of potentiated phospho-protein networks in cancer cells." Cell **118**(2): 217-228.

Irish, J. M., J. H. Myklebust, A. A. Alizadeh, R. Houot, J. P. Sharman, D. K. Czerwinski, G. P. Nolan and R. Levy (2010). "B-cell signaling networks reveal a negative prognostic human lymphoma cell subset that emerges during tumor progression." Proc Natl Acad Sci U S A **107**(29): 12747-12754.

Irish, J. M., J. H. Myklebust, A. A. Alizadeh, R. Houot, J. P. Sharman, D. K. Czerwinski, G. P. Nolan and R. Levy (2010). "B-cell signaling networks reveal a negative prognostic human lymphoma cell subset that emerges during tumor progression." Proceedings of the National Academy of Sciences of the United States of America **107**(29): 12747-12754.

Kalina, T., J. Flores-Montero, V. H. van der Velden, M. Martin-Ayuso, S. Bottcher, M. Ritgen, J. Almeida, L. Lhermitte, V. Asnafi, A. Mendonca, R. de Tute, M. Cullen, L. Sedek, M. B. Vidriales, J. J. Perez, J. G. te Marvelde, E. Mejstrikova, O. Hrusak, T. Szczepanski, J. J. van Dongen, A. Orfao and C. EuroFlow (2012). "EuroFlow standardization of flow cytometer instrument settings and immunophenotyping protocols." Leukemia **26**(9): 1986-2010.

Kim, D., V. S. Donnenberg, J. W. Wilson and A. D. Donnenberg (2016). "The use of simultaneous confidence bands for comparison of single parameter fluorescent intensity data." Cytometry A **89**(1): 89-97.

Klco, J. M., D. H. Spencer, C. A. Miller, M. Griffith, T. L. Lamprecht, M. O'Laughlin, C. Fronick, V. Magrini, R. T. Demeter, R. S. Fulton, W. C. Eades, D. C. Link, T. A. Graubert, M. J. Walter, E. R. Mardis, J. F. Dipersio, R. K. Wilson and T. J. Ley (2014). "Functional heterogeneity of genetically defined subclones in acute myeloid leukemia." Cancer Cell **25**(3): 379-392.

Kotecha, N., N. J. Flores, J. M. Irish, E. F. Simonds, D. S. Sakai, S. Archambeault, E. Diaz-Flores, M. Coram, K. M. Shannon, G. P. Nolan and M. L. Loh (2008). "Single-cell profiling identifies aberrant STAT5 activation in myeloid malignancies with specific clinical and biologic correlates." Cancer Cell **14**(4): 335-343.

Kotecha, N., P. O. Krutzik and J. M. Irish (2010). "Web-based analysis and publication of flow cytometry experiments." Curr Protoc Cytom **Chapter 10**: Unit10 17.

Krutzik, P., J. Irish, G. Nolan and O. Perez (2004). "Analysis of protein phosphorylation and cellular signaling events by flow cytometry: techniques and clinical applications." Clinical Immunology **110**(3): 206-221.

Krutzik, P. O., M. R. Clutter and G. P. Nolan (2005). "Coordinate analysis of murine immune cell surface markers and intracellular phosphoproteins by flow cytometry." J Immunol **175**(4): 2357-2365.

Landau, D. A., S. L. Carter, P. Stojanov, A. McKenna, K. Stevenson, M. S. Lawrence, C. Sougnez, C. Stewart, A. Sivachenko, L. Wang, Y. Wan, W. Zhang, S. A. Shukla, A. Vartanov, S. M. Fernandes, G. Saksena, K. Cibulskis, B. Tesar, S. Gabriel, N. Hacohen, M. Meyerson, E. S. Lander, D. Neuberg, J. R. Brown, G. Getz and C. J. Wu (2013). "Evolution and impact of subclonal mutations in chronic lymphocytic leukemia." Cell **152**(4): 714-726.

Lee, J. A., J. Spidlen, K. Boyce, J. Cai, N. Crosbie, M. Dalphin, J. Furlong, M. Gasparetto, M. Goldberg, E. M. Goralczyk, B. Hyun, K. Jansen, T. Kollmann, M. Kong, R. Leif, S. McWeeney, T. D. Moloshok, W. Moore, G. Nolan, J. Nolan, J. Nikolich-Zugich, D. Parrish, B. Purcell, Y. Qian, B. Selvaraj, C. Smith, O. Tchuvatkina, A. Wertheimer, P. Wilkinson, C. Wilson, J. Wood, R. Zigon, F. International Society for Advancement of Cytometry Data Standards Task, R. H. Scheuermann and R. R. Brinkman (2008). "MIFlowCyt: the minimum information about a Flow Cytometry Experiment." Cytometry A **73**(10): 926-930.

Leelatian, N., K. E. Diggins and J. M. Irish (2015). Characterizing Phenotypes and Signaling Networks of Single Human Cells by Mass Cytometry. Methods in Molecular Biology**:** 99--113.

Leelatian, N., D. B. Doxie, A. R. Greenplate, B. C. Mobley, J. M. Lehman, J. Sinnaeve, R. M. Kauffmann, J. A. Werkhaven, A. M. Mistry, K. D. Weaver, R. C. Thompson, P. P. Massion, M. A. Hooks, M. C. Kelley, L. B. Chambless, R. A. Ihrie and J. M. Irish (2016). "Single cell analysis of human tissues and solid tumors with mass cytometry." Cytometry B Clin Cytom.

Levine, J. H., E. F. Simonds, S. C. Bendall, K. L. Davis, A. D. Amir el, M. D. Tadmor, O. Litvin, H. G. Fienberg, A. Jager, E. R. Zunder, R. Finck, A. L. Gedman, I. Radtke, J. R. Downing, D. Pe'er and G. P. Nolan (2015). "Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis." Cell **162**(1): 184-197.

Linderman M, Q. P., Simonds E and Bjornson Z. "spade: SPADE - An analysis and visualization tool for Flow Cytometry." R package version 1.14.0 http://cytospade.org.

Lo K, H. F., Brinkman R and Gottardo R (2009). "FlowClust: a Bioconductor package for automated gating of flow cytometry data." BMC Bioinformatics **10**(R package version 3.1.0.).

Loke, P. and J. P. Allison (2003). "PD-L1 and PD-L2 are differentially regulated by Th1 and Th2 cells." Proc Natl Acad Sci U S A **100**(9): 5336-5341.

Maecker, H. T., A. Rinfret, P. D'Souza, J. Darden, E. Roig, C. Landry, P. Hayes, J. Birungi, O. Anzala, M. Garcia, A. Harari, I. Frank, R. Baydo, M. Baker, J. Holbrook, J. Ottinger, L. Lamoreaux, C. L. Epling, E. Sinclair, M. A. Suni, K. Punt, S. Calarota, S. El-Bahi, G. Alter, H. Maila, E. Kuta, J. Cox, C. Gray, M. Altfeld, N. Nougarede, J. Boyer, L. Tussey, T. Tobery, B. Bredt, M. Roederer, R. Koup, V. C. Maino, K. Weinhold, G. Pantaleo, J. Gilmour, H. Horton and R. P. Sekaly (2005). "Standardization of cytokine flow cytometry assays." BMC Immunol **6**: 13.

Mahnke, Y., P. Chattopadhyay and M. Roederer (2010). "Publication of optimized multicolor immunofluorescence panels." Cytometry A **77**(9): 814-818.

Maley, C. C., P. C. Galipeau, J. C. Finley, V. J. Wongsurawat, X. Li, C. A. Sanchez, T. G. Paulson, P. L. Blount, R. A. Risques, P. S. Rabinovitch and B. J. Reid (2006). "Genetic clonal diversity predicts progression to esophageal adenocarcinoma." Nat Genet **38**(4): 468-473.

Marusyk, A., D. P. Tabassum, P. M. Altrock, V. Almendro, F. Michor and K. Polyak (2014). "Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity." Nature **514**(7520): 54-58.

Mason, K. D., S. K. Juneja and J. Szer (2006). "The immunophenotype of acute myeloid leukemia: is there a relationship with prognosis?" Blood Rev **20**(2): 71-82.

McCarthy, N. (2013). "Heterogeneity: A multidimensional overview." Nat Rev Cancer **13**(7): 439.

Meehan, S., G. Walther, W. Moore, D. Orlova, C. Meehan, D. Parks, E. Ghosn, M. Philips, E. Mitsunaga, J. Waters, A. Kantor, R. Okamura, S. Owumi, Y. Yang, L. A. Herzenberg and L. A. Herzenberg (2014). "AutoGate: automating analysis of flow cytometry data." Immunol Res **58**(2-3): 218-223.

Moore, W. A. and D. R. Parks (2012). "Update for the logicle data scale including operational code implementations." Cytometry Part A **81A**(4): 273-277.

Mosmann, T. R., I. Naim, J. Rebhahn, S. Datta, J. S. Cavenaugh, J. M. Weaver and G. Sharma (2014). "SWIFT-scalable clustering for automated identification of rare cell populations in large, high-dimensional flow cytometry datasets, Part 2: Biological evaluation." Cytometry A.

Mroz, E. A., A. D. Tward, C. R. Pickering, J. N. Myers, R. L. Ferris and J. W. Rocco (2013). "High intratumor genetic heterogeneity is related to worse outcome in patients with head and neck squamous cell carcinoma." Cancer **119**(16): 3034-3042.

Naim, I., S. Datta, J. Rebhahn, J. S. Cavenaugh, T. R. Mosmann and G. Sharma (2014). "SWIFT-scalable clustering for automated identification of rare cell populations in large, high-dimensional flow cytometry datasets, Part 1: Algorithm design." Cytometry A.

Nervi, B., P. Ramirez, M. P. Rettig, G. L. Uy, M. S. Holt, J. K. Ritchey, J. L. Prior, D. Piwnica-Worms, G. Bridger, T. J. Ley and J. F. DiPersio (2009). "Chemosensitization of acute myeloid leukemia (AML) following mobilization by the CXCR4 antagonist AMD3100." Blood **113**(24): 6206-6214.

Newell, E. W., N. Sigal, S. C. Bendall, G. P. Nolan and M. M. Davis (2012). "Cytometry by time-of-flight shows combinatorial cytokine expression and virus-specific cell niches within a continuum of CD8+ T cell phenotypes." Immunity **36**(1): 142-152.

Nicholas, K. J., A. R. Greenplate, D. K. Flaherty, B. K. Matlock, J. S. Juan, R. M. Smith, J. M. Irish and S. A. Kalams (2015). "Multiparameter analysis of stimulated human peripheral blood mononuclear cells: A comparison of mass and fluorescence cytometry." Cytometry A.

Ornatsky, O., D. Bandura, V. Baranov, M. Nitz, M. A. Winnik and S. Tanner (2010). "Highly multiparametric analysis by mass cytometry." J Immunol Methods **361**(1-2): 1-20.

Ornatsky, O., V. I. Baranov, D. R. Bandura, S. D. Tanner and J. Dick (2006). "Multiple cellular antigen detection by ICP-MS." J Immunol Methods **308**(1-2): 68-76.

Ornatsky, O. I., X. Lou, M. Nitz, S. Schafer, W. S. Sheldrick, V. I. Baranov, D. R. Bandura and S. D. Tanner (2008). "Study of cell antigens and intracellular DNA by identification of element-containing labels and metallointercalators using inductively coupled plasma mass spectrometry." Anal Chem **80**(7): 2539-2547.

Overton, W. R. (1988). "Modified histogram subtraction technique for analysis of flow cytometry data." Cytometry **9**(6): 619-626.

Parks, D., M. Roederer and W. A. Moore (2004). ""Logicle" functions provide improved displays of FACS data and avoid the deceptive effects of log scaling for low signals." Cytometry Part A **59A**(1): 87-87.

Patel, A. P., I. Tirosh, J. J. Trombetta, A. K. Shalek, S. M. Gillespie, H. Wakimoto, D. P. Cahill, B. V. Nahed, W. T. Curry, R. L. Martuza, D. N. Louis, O. Rozenblatt-Rosen, M. L. Suva, A. Regev and B. E. Bernstein (2014). "Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma." Science **344**(6190): 1396-1401.

Patel, J. P., M. Gonen, M. E. Figueroa, H. Fernandez, Z. Sun, J. Racevskis, P. Van Vlierberghe, I. Dolgalev, S. Thomas, O. Aminova, K. Huberman, J. Cheng, A. Viale, N. D. Socci, A. Heguy, A. Cherry, G. Vance, R. R. Higgins, R. P. Ketterling, R. E. Gallagher, M. Litzow, M. R. van den Brink, H. M. Lazarus, J. M. Rowe, S. Luger, A. Ferrando, E. Paietta, M. S. Tallman, A. Melnick, O. Abdel-Wahab and R. L. Levine (2012). "Prognostic relevance of integrated genetic profiling in acute myeloid leukemia." N Engl J Med **366**(12): 1079-1089.

Polikowsky, H. G., C. E. Wogsland, K. E. Diggins, K. Huse and J. M. Irish (2015). "Cutting Edge: Redox Signaling Hypersensitivity Distinguishes Human Germinal Center B Cells." J Immunol **195**(4): 1364-1367.

Prince, M. E., R. Sivanandan, A. Kaczorowski, G. T. Wolf, M. J. Kaplan, P. Dalerba, I. L. Weissman, M. F. Clarke and L. E. Ailles (2007). "Identification of a subpopulation of cells with cancer stem cell properties in head and neck squamous cell carcinoma." Proc Natl Acad Sci U S A **104**(3): 973-978.

Pyne, S., X. Hu, K. Wang, E. Rossin, T. I. Lin, L. M. Maier, C. Baecher-Allan, G. J. McLachlan, P. Tamayo, D. A. Hafler, P. L. De Jager and J. P. Mesirov (2009). "Automated high-dimensional flow cytometric data analysis." Proc Natl Acad Sci U S A **106**(21): 8519-8524.

Pyne, S., S. X. Lee, K. Wang, J. Irish, P. Tamayo, M. D. Nazaire, T. Duong, S. K. Ng, D. Hafler, R. Levy, G. P. Nolan, J. Mesirov and G. J. McLachlan (2014). "Joint modeling and registration of cell populations in cohorts of high-dimensional flow cytometric data." PLoS One **9**(7): e100334.

Qian, Y., C. Wei, F. Eun-Hyung Lee, J. Campbell, J. Halliley, J. A. Lee, J. Cai, Y. M. Kong, E. Sadat, E. Thomson, P. Dunn, A. C. Seegmiller, N. J. Karandikar, C. M. Tipton, T. Mosmann, I. Sanz and R. H. Scheuermann (2010). "Elucidation of seventeen human peripheral blood B-cell subsets and quantification of the tetanus response using a density-based method for the automated identification of cell populations in multidimensional flow cytometry data." Cytometry B Clin Cytom **78 Suppl 1**: S69-82.

Qiu, P., E. F. Simonds, S. C. Bendall, K. D. Gibbs, Jr., R. V. Bruggner, M. D. Linderman, K. Sachs, G. P. Nolan and S. K. Plevritis (2011). "Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE." Nat Biotechnol **29**(10): 886-891.

Roederer, M., W. Moore, A. Treister, R. R. Hardy and L. A. Herzenberg (2001). "Probability binning comparison: a metric for quantitating multivariate distribution differences." Cytometry **45**(1): 47-55.

Roweis, S. T. and L. K. Saul (2000). "Nonlinear dimensionality reduction by locally linear embedding." Science **290**(5500): 2323-2326.

Sarry, J. E., K. Murphy, R. Perry, P. V. Sanchez, A. Secreto, C. Keefer, C. R. Swider, A. C. Strzelecki, C. Cavelier, C. Recher, V. Mansat-De Mas, E. Delabesse, G. Danet-Desnoyers and M. Carroll (2011). "Human acute myelogenous leukemia stem cells are rare and heterogeneous when assayed in NOD/SCID/IL2Rgammac-deficient mice." J Clin Invest **121**(1): 384-395.

Shah, A. T., K. E. Diggins, A. J. Walsh, J. M. Irish and M. C. Skala (2015). "In Vivo Autofluorescence Imaging of Tumor Heterogeneity in Response to Treatment." Neoplasia **17**(12): 862-870.

Shekhar, K., P. Brodin, M. M. Davis and A. K. Chakraborty (2013). "Automatic Classification of Cellular Expression by Nonlinear Stochastic Embedding (ACCENSE)." Proc Natl Acad Sci U S A.

Shekhar, K., P. Brodin, M. M. Davis and A. K. Chakraborty (2014). "Automatic Classification of Cellular Expression by Nonlinear Stochastic Embedding (ACCENSE)." Proc Natl Acad Sci U S A **111**(1): 202-207.

Smyth, M. J., S. F. Ngiow, A. Ribas and M. W. Teng (2016). "Combination cancer immunotherapies tailored to the tumour microenvironment." Nat Rev Clin Oncol **13**(3): 143-158.

Spidlen, J., A. Barsky, K. Breuer, P. Carr, M. D. Nazaire, B. A. Hill, Y. Qian, T. Liefeld, M. Reich, J. P. Mesirov, P. Wilkinson, R. H. Scheuermann, R. P. Sekaly and R. R. Brinkman (2013). "GenePattern flow cytometry suite." Source Code Biol Med **8**(1): 14.

Spidlen, J., C. Bray, I. D. S. T. Force and R. R. Brinkman (2015). "ISAC's classification results file format." Cytometry A **87**(1): 86-88.

Spidlen, J., K. Breuer and R. Brinkman (2012). "Preparing a Minimum Information about a Flow Cytometry Experiment (MIFlowCyt) compliant manuscript using the International Society for Advancement of Cytometry (ISAC) FCS file repository (FlowRepository.org)." Curr Protoc Cytom **Chapter 10**: Unit 10 18.

Spidlen, J., R. C. Leif, W. Moore, M. Roederer, F. International Society for the Advancement of Cytometry Data Standards Task and R. R. Brinkman (2008). "Gating-ML: XML-based gating descriptions in flow cytometry." Cytometry A **73A**(12): 1151-1157.

Spidlen, J., W. Moore, D. Parks, M. Goldberg, C. Bray, P. Bierre, P. Gorombey, B. Hyun, M. Hubbard, S. Lange, R. Lefebvre, R. Leif, D. Novo, L. Ostruszka, A. Treister, J. Wood, R. F. Murphy, M. Roederer, D. Sudar, R. Zigon and R. R. Brinkman (2010). "Data File Standard for Flow Cytometry, version FCS 3.1." Cytometry A **77**(1): 97-100.

Spitzer, M. H., P. F. Gherardini, G. K. Fragiadakis, N. Bhattacharya, R. T. Yuan, A. N. Hotson, R. Finck, Y. Carmi, E. R. Zunder, W. J. Fantl, S. C. Bendall, E. G. Engleman and G. P. Nolan (2015). "IMMUNOLOGY. An interactive reference framework for modeling a dynamic immune system." Science **349**(6244): 1259425.

Spoo, A. C., M. Lubbert, W. G. Wierda and J. A. Burger (2007). "CXCR4 is a prognostic marker in acute myelogenous leukemia." Blood **109**(2): 786-791.

Stein, E. M. and M. S. Tallman (2012). "Remission induction in acute myeloid leukemia." Int J Hematol **96**(2): 164-170.

Stetler-Stevenson, M., B. Davis, B. Wood and R. Braylan (2007). "2006 Bethesda International Consensus Conference on Flow Cytometric Immunophenotyping of Hematolymphoid Neoplasia." Cytometry B Clin Cytom **72 Suppl 1**: S3.

Sugar, I. P. and S. C. Sealfon (2010). "Misty Mountain clustering: application to fast unsupervised flow cytometry gating." BMC Bioinformatics **11**: 502.

Supernat, A., S. Lapinska-Szumczyk, H. Majewska, J. Gulczynski, W. Biernat, D. Wydra and A. J. Zaczek (2014). "Tumor Heterogeneity at Protein Level as an Independent Prognostic Factor in Endometrial Cancer." Transl Oncol.

Taussig, D. C., F. Miraki-Moud, F. Anjos-Afonso, D. J. Pearce, K. Allen, C. Ridler, D. Lillington, H. Oakervee, J. Cavenagh, S. G. Agrawal, T. A. Lister, J. G. Gribben and D. Bonnet (2008). "Anti-CD38 antibody-mediated clearance of human repopulating cells masks the heterogeneity of leukemia-initiating cells." Blood **112**(3): 568-575.

Taussig, D. C., J. Vargaftig, F. Miraki-Moud, E. Griessinger, K. Sharrock, T. Luke, D. Lillington, H. Oakervee, J. Cavenagh, S. G. Agrawal, T. A. Lister, J. G. Gribben and D. Bonnet (2010). "Leukemia-initiating cells from some acute myeloid leukemia patients with mutated nucleophosmin reside in the CD34(-) fraction." Blood **115**(10): 1976-1984.

Team, R. C. (2013). "R: A language and environment for statistical computing." R Foundation for Statistical Computing, Vienna, Austria.(http://www.R-project.org/).

Tenenbaum, J. B., V. de Silva and J. C. Langford (2000). "A global geometric framework for nonlinear dimensionality reduction." Science **290**(5500): 2319-2323.

Tong, D. L., G. R. Ball and A. G. Pockley (2015). "gEM/GANN: A multivariate computational strategy for auto-characterizing relationships between cellular and clinical phenotypes and predicting disease progression time using high-dimensional flow cytometry data." Cytometry A.

Valk, P. J., R. G. Verhaak, M. A. Beijen, C. A. Erpelinck, S. B. v. W. van Doorn-Khosrovani, J. M. Boer, H. B. Beverloo, M. J. Moorhouse, P. J. van der Spek and B. Löwenberg (2004). "Prognostically useful gene-expression profiles in acute myeloid leukemia." New England Journal of Medicine **350**(16): 1617-1628.

van der Maaten, L. and G. Hinton (2008). "Visualizing Data using t-SNE." Journal of Machine Learning Research **9**: 2579-2605.

van Dongen, J. J., L. Lhermitte, S. Bottcher, J. Almeida, V. H. van der Velden, J. Flores-Montero, A. Rawstron, V. Asnafi, Q. Lecrevisse, P. Lucio, E. Mejstrikova, T. Szczepanski, T. Kalina, R. de Tute, M. Bruggemann, L. Sedek, M. Cullen, A. W. Langerak, A. Mendonca, E. Macintyre, M. Martin-Ayuso, O. Hrusak, M. B. Vidriales, A. Orfao and C. EuroFlow (2012). "EuroFlow antibody panels for standardized n-dimensional flow cytometric immunophenotyping of normal, reactive and malignant leukocytes." Leukemia **26**(9): 1908-1975.

van Dongen, J. J., A. Orfao and C. EuroFlow (2012). "EuroFlow: Resetting leukemia and lymphoma immunophenotyping. Basis for companion diagnostics and personalized medicine." Leukemia **26**(9): 1899-1907.

Van Gassen, S., B. Callebaut, M. J. Van Helden, B. N. Lambrecht, P. Demeester, T. Dhaene and Y. Saeys (2015). "FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data." Cytometry A.

Vogel, C. and E. M. Marcotte (2012). "Insights into the regulation of protein abundance from proteomic and transcriptomic analyses." Nat Rev Genet **13**(4): 227-232.

Walter, M. J., D. Shen, L. Ding, J. Shao, D. C. Koboldt, K. Chen, D. E. Larson, M. D. McLellan, D. Dooling and R. Abbott (2012). "Clonal architecture of secondary acute myeloid leukemia." New England Journal of Medicine **366**(12): 1090-1098.

Wong, M. T., D. E. Ong, F. S. Lim, K. W. Teng, N. McGovern, S. Narayanan, W. Q. Ho, D. Cerny, H. K. Tan, R. Anicete, B. K. Tan, T. K. Lim, C. Y. Chan, P. C. Cheow, S. Y. Lee, A. Takano, E. H. Tan, J. K. Tam, E. Y. Tan, J. K. Chan, K. Fink, A. Bertoletti, F. Ginhoux, M. A. Curotto de Lafaille and E. W. Newell (2016). "A High-Dimensional Atlas of Human T Cell Diversity Reveals Tissue-Specific Trafficking and Cytokine Signatures." Immunity **45**(2): 442-456.

Wood, B. L., M. Arroz, D. Barnett, J. DiGiuseppe, B. Greig, S. J. Kussick, T. Oldaker, M. Shenkin, E. Stone and P. Wallace (2007). "2006 Bethesda International Consensus recommendations on the immunophenotypic analysis of hematolymphoid neoplasia by flow cytometry: optimal reagents and reporting for the flow cytometric diagnosis of hematopoietic neoplasia." Cytometry B Clin Cytom **72 Suppl 1**: S14-22.

Zare, H., P. Shooshtari, A. Gupta and R. R. Brinkman (2010). "Data reduction for spectral clustering to analyze high throughput flow cytometry data." BMC Bioinformatics **11**: 403.

Zare, H., J. Wang, A. Hu, K. Weber, J. Smith, D. Nickerson, C. Song, D. Witten, C. A. Blau and W. S. Noble (2014). "Inferring clonal composition from multiple sections of a breast cancer." PLoS Comput Biol **10**(7): e1003703.