Seminars in Immunology xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Seminars in Immunology

journal homepage: www.elsevier.com/locate/ysmim

Review

Each cell counts: Hematopoiesis and immunity research in the era of single cell genomics

Diego Adhemar Jaitin[a], Hadas Keren-Shaul[a], Naama Elefant[a,b,c], Ido Amit[a,*]

[a] Department of Immunology, Weizmann Institute of Science, Rehovot, Israel
[b] Department of Family Medicine, Clalit Health Services, Jerusalem, Israel
[c] The Hebrew University, Jerusalem, Israel

ARTICLE INFO

Keywords:
Single-cell genomics
Transcriptional profiling
Hematopoiesis
Single-cell transcriptomics
Immunity
Personalized medicine

ABSTRACT

Hematopoiesis and immunity are mediated through complex interactions between multiple cell types and states. This complexity is currently addressed following a reductionist approach of characterizing cell types by a small number of cell surface molecular features and gross functions. While the introduction of global transcriptional profiling technologies enabled a more comprehensive view, heterogeneity within sampled populations remained unaddressed, obscuring the true picture of hematopoiesis and immune system function. A critical mass of technological advances in molecular biology and genomics has enabled genome-wide measurements of single cells – the fundamental unit of immunity. These new advances are expected to boost detection of less frequent cell types and fuzzy intermediate cell states, greatly expanding the resolution of current available classifications. This new era of single-cell genomics in immunology research holds great promise for further understanding of the mechanisms and circuits regulating hematopoiesis and immunity in both health and disease. In the near future, the accuracy of single-cell genomics will ultimately enable precise diagnostics and treatment of multiple hematopoietic and immune related diseases.

1. Introduction

The immune system is a highly complex network of organs, cells and molecules. About two centuries ago, pioneering studies by Edward Jenner and Louis Pasteur brought to realization that immunity is an acquired feature, a concept strengthened with the discovery of humoral immunity by Behring and Kitasato [1]. Around the same time, concepts and theories elaborated by Rudolf Virchow and Elie Metchnikoff among others set the foundation for a cellular theory of immunity [2,3]. Ever since, research in hematopoiesis and immunity involved the characterization of the different immune cell types and their roles in homeostasis and defense against pathogens. From the outset, immune cells have been characterized by their apparent function, with the first observations made using gross phenotypes (size, shape, adherence to surfaces, etc.) and cell staining dyes.

Significant breakthroughs in our understanding of cellular and molecular immunity arose 30 years ago with advances in single-cell technologies, catapulted by efficient production of monoclonal antibodies and the invention of fluorescence-activated cell sorting (FACS). These technologies enabled the classification of immune cells based on extracellular molecular features and are still the state of the art in immune and medical research [4,5]. However, multiple recent studies demonstrate that the complexity of cell states and types cannot be resolved by these surface markers, which in many cases have only a vague association to cell function/identity [6–9]. Therefore the precise definition of many immune cell types remains controversial [6–8,10,11]. Importantly, the lack of resolution for discerning the heterogeneity of cell types in complex tissues in an unbiased way, rendered the etiology of several diseases obscure [12]. Effective treatment for inflammatory driven diseases such as rheumatoid arthritis (RA) and multiple sclerosis (MS) is still lacking [13]. Indeed, after many years of intensive research their definite pathology is still unknown [12]. This is in part due to the limited resolution current technologies have for the required distinctions between cell types and cell states, rendering a rather poor per-patient characterization of the immune cells and pathways involved. The prospect for addressing this complexity on a genome-wide single-cell level holds promise for gaining new insights not only into basic function of the immune system, but also regarding the etiology of many of these conditions. In this review, we introduce the emerging field of single-cell transcriptomics as a new tool

* Corresponding author.
E-mail address: ido.amit@weizmann.ac.il (I. Amit).

http://dx.doi.org/10.1016/j.smim.2015.01.002
1044-5323/

to better understand the cells and pathways of hematopoiesis and immune response, and the roles they play in immunopathology.

2. The single-cell RNA sequencing revolution

For a long time, immunology has been studied by averaging cellular phenotypes to draw conclusions on how the individual cell components behave, masking the real distribution of functions and variability in gene expression between individual cells. Consequently, the individual cell behavior could have only been extrapolated from the average behavior of a population of cells [14]. However, the expression profile from a mass of cells does not represent the expression profile of each individual one. Cells demonstrate dynamic behaviors within the same organ or even in the same sub-region of an organ. A particular cell type may represent a plethora of sub-behaviors, such as the expression of various chemokine/growth receptors and inflammatory molecules under different conditions. Hence, a major bottleneck for biological research is a global in vivo characterization of individual cell states within complex tissues [15–17]. To address this, there has recently been an exciting surge of new technologies aimed at facilitating genome-wide single-cell measurements [18]. This review will focus on advances in single-cell RNA-seq approaches and how they may impact hematopoiesis and immune research.

Methods for single-cell genome-wide expression analysis are continuously being developed, with increasing coverage, precision and throughput. The first genome-wide transcription analysis of single cells was done using microarrays [19], demonstrating the potential of single-cell genome-wide approaches; however, expression arrays are laborious, expensive and relatively inaccurate, limiting the utility of single-cell expression array for research. A revolution in the field occurred with the application of next-generation sequencing for transcriptome analysis [20,21]. The first reported single-cell RNA sequencing technology was a variation of the single-cell microarray technique adopted for next-generation sequencing and was based on poly-A extension at the 3′ end of the cDNA molecule [22,23]. This work was soon followed by multiple single-cell methods that improved throughput, for example single-cell tagged reverse transcription (STRT) [24], and coverage, such as template-switching mechanism at the end of the cDNA (SMART) [25,26]. Other protocols replaced the initial pre-amplification PCR step by introducing in vitro transcription (Cell-Seq) [27]. In general, the methods for generating single cell RNA sequencing libraries can be broken down into two main approaches: full transcriptome coverage [22,26,28], and molecular counting methods that are based on sequencing of the 5′ or the 3′ end of the molecule [24,27,29]. Counting methods further evolved by addition of unique molecular identifiers (UMI), using random barcodes during the cDNA synthesis, an important feature that enables counting unique RNA molecules. Addition of UMI and relevant analytical methods resolved many biases associated with single-cell RNA-seq analysis [29–31].

Some of the difficulties associated with single-cell RNA-seq include accurate capturing of single cells, which can be challenging depending on the tissue of choice and the specific method used. Microfluidics-based devices such as the C1 from Fluidigm are currently used as the leading commercial single-cell capture approach [32–34]. The lower volumes of microfluidics produce expression libraries that recover single-cell gene expression with a slightly higher coverage, albeit allowing the processing of only a few tens of cells per experiment, rendering this system less cost-effective and impractical for comprehensive studies [32–34] Consequently, a critical factor that makes in vivo single-cell analysis relevant is the number of cells that can be analyzed in parallel in a reasonable amount of time [35]. Indeed, a major challenge for implementation of single-cell RNA-seq routinely in research is the ability to measure tens of thousands of single cells in an affordable manner. We developed massively parallel single-cell RNA-seq (MARS-seq) to overcome many of the aforementioned hurdles of single-cell RNA-seq research. The MARS-seq pipeline is coupled to FACS, enabling sorting of individual cells into 384-well plates and a streamlined and automated molecular barcoding and amplification protocol to obtain thousands of single cell RNA-seq libraries a day (Fig. 1) [29]. Another important feature of MARS-seq is the ability to combine thousands of single cells on a single sequencing flow cell [29], reducing the total cost to less then 50 cents per cell.

Analysis of tissue heterogeneity requires surveying thousands of cells to ensure robust sampling of rare cell types and cellular states [36]. From our experience, orders of magnitude in abundance between cell types and states are common to most tissues. Thus, a thousand cells is the minimum number of cells required to provide the necessary power to robustly detect rare cell populations and states. Such is the case when studying bone marrow hematopoietic progenitors [37], or cell composition and states during an immune response [29]. Importantly, biological interpretation of single-cell expression data is not trivial and critically depends on emerging computational approaches to accurately quantify gene expression, while accounting for technical noise [28,29,31,35,36,38–40]. An emerging challenge is to accurately cluster cells into coherent groups on the basis of their gene expression profiles. These computational methods are expected to constantly develop and improve our ability to extract deep biological insights from the rich data emerging from single-cell genome-wide experiments.

3. Single-cell expression analysis in the immune system

Complex tissues of the immune system, such as the bone marrow and the lymph node, are currently studied assuming a predefined cell composition. Cells in these organs have been previously identified in a top-down fashion [41,42], where cells are first identified by function or phenotype, mapping this function into a fixed position in the hematopoietic tree. This approach, albeit effective, may lead to biases because markers describe a heterogeneous population [42].

3.1. Heterogeneity within 'homogeneous' cell populations

One case that illustrates the circular logic in cell definition is the seminal discovery of the dendritic cell (DC) by Steinman [41], and the later identification of a DC marker, integrin alpha X (CD11c) [9], which has since been used to enrich for DCs [11]. DCs are antigen-presenting cells that were originally characterized through their unique morphology [41], but are now understood to represent a highly heterogeneous group [43] with multiple functions, regulatory circuits and phenotypes [7,8,41]. Despite considerable efforts and progress using the marker-based approach to sub-classify this group, much of the known functional heterogeneity within the DC compartment is not truly compatible with any of the DC sub-classification schemes [7,8,44]. Such lack of definitive models for cell types and states is paradigmatic of many other fields in biology. Single-cell RNA-seq provides an alternative and comprehensive approach to define cell types and functional cell states (Fig. 1) [29,42,45]. An unbiased map of cells sampled directly from tissues may revolutionize our ability to understand their function in immune responses and diseases. Additionally, single-cell techniques will also provide valuable information on the stochastic nature of gene regulation and the genetic programs orchestrating the differentiation and function of the different immune cells, potentially revising the models of their subtypes, lineages and their compositions within tissues. Furthermore, such data should greatly

Please cite this article in press as: Jaitin DA, et al. Each cell counts: Hematopoiesis and immunity research in the era of single cell genomics. Semin Immunol (2015), http://dx.doi.org/10.1016/j.smim.2015.01.002