Single-Cell Genomics in Immunology

Introduction

Why single-cell analysis?

Study of the immune system has long relied on single-cell techniques such as flow cytometry and microscopy to explore and define new immune cell subsets and functionalities. Applying single-cell genomics techniques to the field of immunology has allowed deeper profiling of immune cell phenotypes and functional states than bulk-cell analysis and has advanced the field tremendously in recent years. The practical application of these discoveries in the fields of oncology, vaccine research and autoimmunity has led to unprecedented advances in the treatment of human disease. This compendium gives an overview of papers applying single-cell genomics strategies on the C1™ system to unravel the complexities of the human immune system in four key areas:

- Pairing phenotypic profiles with clonal identity via T cell receptor sequences
- Mapping cell transition states and cell fate decisions
- Discovering new cell types and functions
- Capturing cell-specific variations in isoform expression

Targeted versus whole transcriptome analysis

Gene expression profiling at the single-cell level is possible with both targeted and whole transcriptome approaches, such as qPCR and mRNA sequencing (mRNA-seq). The sensitivity of targeted gene expression analysis is particularly useful for reliable detection of genes with characteristically low expression, such as transcription factors, or when analyzing cell types with low overall gene expression, such as naïve lymphocytes. However, targeted gene expression profiling is constrained because it requires *a priori* knowledge of the relevant genes. Single-cell mRNA sequencing is a useful counterpart to targeted gene expression, especially in the field of immunology. Whole transcriptome datasets allow for unbiased assessment of gene expression, and when full-length transcript amplification chemistry is used, these datasets can be mined for expression of the highly diverse gene isoforms that are common to immune cell subsets. Thus, targeted and whole transcriptome detection in tandem serve as powerful and complementary analysis tools, as demonstrated in several of the publications reviewed here.

Linking clonal identity with phenotypic immune response

Why is this important?

A T cell clonotype refers to a clonally expanded population of T cells that share the same TCR α and β chain sequences and thus recognize the same antigen when presented in conjunction with a particular major histocompatibility complex. Clonotype identification and enumeration has been used to characterize the immune response across a wide array of medical conditions, including infection, autoimmunity and cancer. The most common form of clonotype analysis identifies only the β chain sequence rather than the full α and β chain TCR sequence pair. The full TCR sequence is required for definitive clonal identification because two different clonotypes may share an α or β chain sequence. The paired α and β chain sequences are also required to identify the cognate antigen or to reengineer the TCR in therapeutic applications.

The combined single-cell resolution of clonotype and associated cell phenotype has implications for many fundamental immunologic research questions. Clonotype identity provides a built-in barcode to map cell fate decisions. Being able to assess the number and profile of clonotypes that expand in an immune response promises to improve our understanding of how the immune system responds to threat and will inform strategies to either augment that response or, in the case of autoimmune dysfunction, inhibit it.

While findings to date are only preliminary, knowing the TCR sequence and the phenotype of the corresponding cells may also provide a path to biomarker identification and treatment strategies. Strategies to augment the immune response include synthetically engineered TCR molecules or antibodies that target the respective antigen in concert with other activating tools such as the immunological activation that CAR T cells provide. Conversely, in the case of autoimmunity, targeted blocking or selective removal of the autoreactive epitope can be implemented through similar engineering of a TCR or antibody, but used in conjunction with immune-inhibiting rather than immune-activating therapies.

Research

Stubbington et al. "T cell fate and clonality inference from single-cell transcriptomes." *Nature Methods* (2016).

Analyzing the whole transcriptome of T cells has been used to further characterize both T cell subtype and function. The TCR α and β chain sequences are contained within the whole transcriptome data if full-length transcript amplification is implemented, as in the C1 mRNA

Seq workflow using SMART-Seq™ chemistry. However, the TCR sequences are not routinely recovered from whole transcriptome data because of the high sequence variability introduced during the recombination of genes within the variable, diversity and joining regions for both the α and β chain sequences (Figure 1). Stubbington et al. devised a companion TCR analysis pipeline, TraCeR, that reconstructs TCR α and β chain sequences, intended to be run in parallel with conventional gene expression analysis (whole transcriptome amplification, or WTA). After the data is generated from TraCeR and WTA, it may then be recombined for single-cell resolution of both clonotype and phenotype.

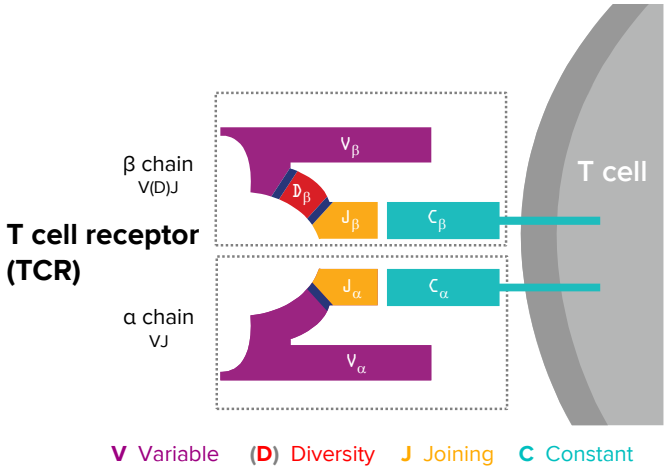

Figure 1: The TCR is comprised of a constant gene region and three variable gene segments, termed variable, diversity and joining. Additional junctional diversity between the variable segments adds complexity to the sequence. The highly variable sequence that is generated after recombination of these gene sets requires an mRNA-seq data analysis pipeline that is separate from conventional gene expression analysis in order to reconstruct α and β chain sequences.

Key takeaways

- The co-localization of the full-length, paired α, β chain TCR sequences with cell phenotype is only possible at the single-cell level because bulk analyses decouple the α, β chain link that identifies a specific clonotype.
- Through use of a single-cell, full-length transcript amplification protocol, such as C1 mRNA Seq using SMART-Seq chemistry, the complexity of the T cell repertoire can be accurately captured and viewed in the context of clonotypically related gene expression profiles.
- The authors have made TraCeR available for the research community through Script Hub™.

Mapping transition states and cell fate decisions

Why is this important?

Immune cell populations arise from precursor cells passing through myriad checkpoints and cell states in immune system tissues including the bone marrow, thymus, lymph nodes and spleen. Dynamic processes such as hematopoiesis, thymic selection and pathogen exposure continually shape and impact immune cell fate decisions and determine cell phenotype and function. Mapping transition states and understanding how cell fate decisions are made provide important insights to fundamental aspects of immune system development and to our understanding of the immune response in health and disease. For example, a cognate antigen encounter may stimulate a naïve T cell to proliferate and differentiate into a variety of subtypes (Figure 2). The activity and state of these proliferated cells, while directed against the cognate antigen, are significantly moderated by interaction between the responding cells and other adjacent cell types. Cell activation is necessary in order to resolve the immune response but can also have adverse physiological consequences or can become pathogenic, creating autoimmune dysfunction.

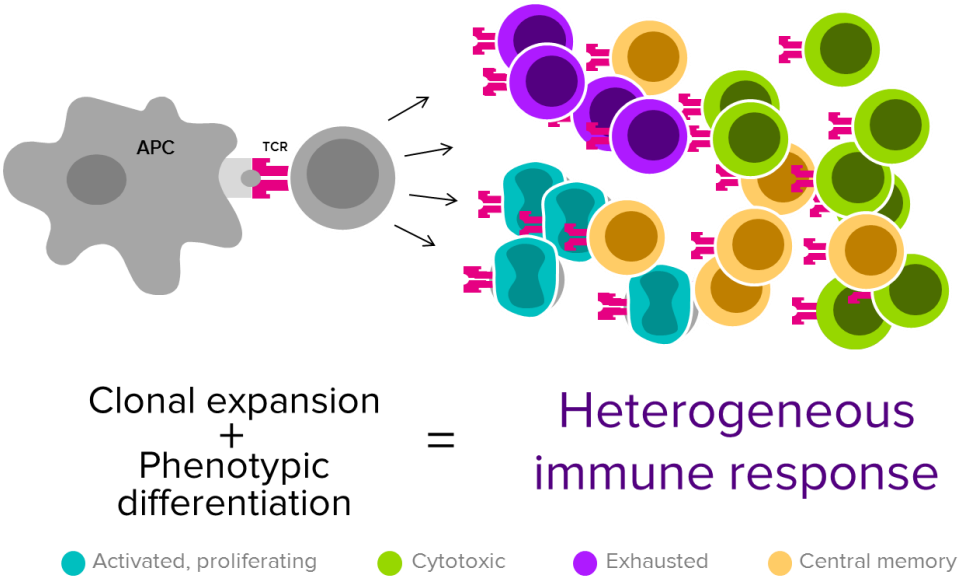

Figure 2. A naïve T cell antigen encounter with an antigen-presenting cell (APC) stimulates a proliferative response in which the naïve T cell expands into a population of various T cell phenotypes that present an aggregate immune response. This response and its cellular composition change with the location in the tissue and organism and with stage of the response. Understanding how, when and where immune cells of various phenotypes and functions arise during an immune response can guide development of therapeutic interventions.

The depth of information available from single-cell studies on the C1 system will drive further refinements in the understanding of immune-cell function and subtype. For example, T helper (CD4+) cells are categorized within an array of different cell types from Th1 and Th2 to Th17, Th9, Th22, regulatory T cells and follicular helper T cells. Further differentiations exist within these categories, and as the understanding of the differentiation and function of

these cells becomes more advanced, the discrete binning of cells is expected to shift into a more spectrum-based phenotype assessment (Figure 3).

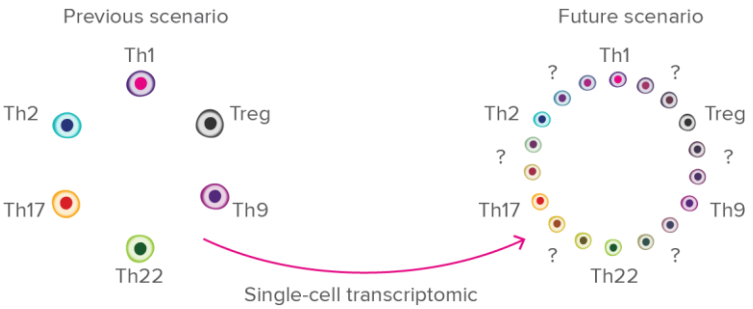

Figure 3. T cell physiology is increasingly understood to encompass phenotype plasticity and diversity of function. As this understanding becomes more comprehensive, discrete T cell phenotypes may shift to a more spectrum-based continuum.

Research

Olsson et al. "Single-cell analysis of mixed-lineage states leading to a binary cell fate choice." *Nature* (2016).

A continuing challenge in developmental biology is delineation of the rare, intermediate cell states and the corresponding regulatory gene networks that drive cell fate decisions. Olsson et al. performed single-cell mRNA-seq on 382 cells of the mouse hematopoietic system spanning nine phenotypically well-defined stages of myeloid differentiation. They identified previously unrecognized mixed-lineage states in the differentiation of neutrophils and macrophages from hematopoietic stem cell progenitors and distinct combinations of transcription factor expression that correlate with cell state. They hypothesize that these mixed-lineage states are dynamically unstable and express counteracting gene regulatory networks that ultimately drive cells to either a monocytic or a granulocytic potential.

Proserpio et al. "Single-cell analysis of CD4+ T-cell differentiation reveals three major cell states and progressive acceleration of proliferation." *Genome Biology* (2016).

Proserpio et al. identified a sequence of three cell states in the path to Th2 differentiation using single-cell mRNA-seq (78 cells), followed by verification with single-cell qPCR (48 cells per subpopulation). The three states of Th2 differentiation are 1) recently naïve cells that have been activated, 2) differentiated cells and 3) cytokine-expressing cells. The cell state of differentiation correlated with a higher proliferation rate compared to that of the activation cell state. The authors hypothesize that the differentiation process reduces heterogeneity in the proliferating population, allowing maximal and homogeneous proliferation of the cell types responsible for effective pathogen clearance. The same path to differentiation was also confirmed in Th1 cells using single-cell mRNA-seq.

Key takeaways

- Single-cell biology is necessary to capture highly transient, and therefore rare, cell states within a larger population of cells.
- Transcription factors, which are often tied to cell fate decisions, frequently have lower expression levels than other transcripts (Vaquerizas et al., 2009) requiring high sensitivity in the single cell genomic approach.
- Functions such as proliferation are tied to specific stages of cell differentiation. Mapping the sequence of cell states can also further refine the understanding of specific cell function.

Discovering new cell types and functions

Why is this important?

The immune system functions through the complex interplay of numerous cell types. While a certain cell type may have a generally characterized role, finer delineations or subroles exist within a cell type and may vary with time or location. The higher resolution enabled by single-cell genomics reveals the true level of cell heterogeneity and previously uncharacterized cell subtypes and aspects of cell function.

Research

Mahata et al. "Single-cell RNA sequencing reveals T helper cells synthesizing steroids de novo to contribute to immune homeostasis." *Cell Reports* (2014).

Mahata el al. identified a new subpopulation of Th2 cells with the C1 system using both targeted qPCR and mRNA-seq. Targeted qPCR of 96 genes per cell was generated with the Biomark™ system, which pairs directly with the single-cell cDNA libraries generated with the C1 system. While C1 combined with Biomark was used for efficient preliminary hypothesis testing, the authors used mRNA-seq as a discovery tool to identify potential surface markers that were used to sort a newly identified subpopulation of immunosuppressive cells for *ex vivo* functional validation. The newly identified Th2 cell type had immunosuppressive activity via production of the steroid hormone pregnenolone, which has been implicated in the inhibition of Th2 cell proliferation and B cell class switching.

Gaublomme et al. "Single-cell genomics unveils critical regulators of Th17 cell pathogenicity." *Cell* (2015).

Gaublomme et al. used single-cell mRNA-seq to characterize the function of Th17 cells (722 in total) in an autoimmune model of multiple sclerosis (MS). The Th17 T cell subtype acts as host defense in infection, although improper recognition can generate an autoimmune response. Th17 cells found in the central nervous system (CNS), the site of autoimmunity in

the MS model, were compared to Th17 cells in the lymph node (LN), a reservoir for systemic, physiologically normal immune cells. The Th17 cells from the CNS, which had become pathogenic, were associated with a set of genes that were differentially expressed compared to the nonpathogenic LN sample. These genes serve as pathogenic markers and may provide a set of targets for therapeutic development.

Key takeaways

- Immune cell subtype categories are being further refined as our understanding of cell roles improves. Through the use of single-cell genomics, cell subsets previously believed to be discrete are found to contain more than one functional cell type, and the characteristic traits of different cell types have been found to overlap.
- Capturing the single-cell transcriptional profiles in an autoimmune model reveals how the cell function changes from physiologically normal to pathogenic and provides targets for therapeutic modification.

Capturing isoform expression differences

Why is this important?

Cellular processes can be modulated by the expression of different transcript, and consequently protein, isoforms. Variations in alternatively spliced exons of the mRNA transcripts can generate functional differences in the resulting proteins (Figure 4). Understanding the relative abundance of a set of isoforms can be used to make inferences on corresponding variations in physiological activity. This is true for certain immunologically associated proteins, including the checkpoint inhibitor CTLA-4 (Ward et al., 2014) and the Treg-associated protein Foxp3 (Joly et al., 2015) for which isoform configurations are tied to autoimmunity. There is also some evidence that protein isoforms change with T cell state, such as with CD28 in naïve and memory T cells (Deshpande et al., 2002). These are just a few examples. As the literature shows, differential isoform expression is widespread and in many cases has largely uncharacterized physiological implications.

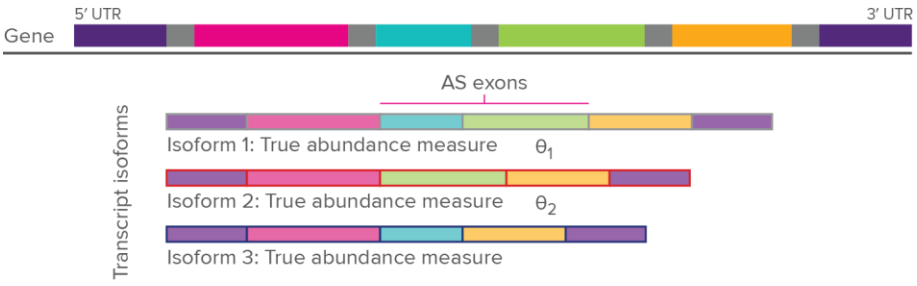

Figure 4. For a given gene, multiple isoforms can exist due to alternative exon splicing, resulting in great variation within the resulting transcripts. Correlating the relative expression levels with patterns in gene expression at the single-cell level provides insight into how specific isoforms are tied to different physiological processes.

Research

Shalek et al. "Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells." *Nature* (2013).

Shalek et al. used C1 mRNA-seq to investigate transcriptional changes in bone marrow-derived dendritic cells (BMDCs) with LPS activation. Because the mRNA-seq data contains full-length transcript datasets, the authors were able to identify a previously unrecognized variation in gene isoform expression, pointing to the need for further study regarding isoform specificity in understanding the immune response. Other variations in the genomic response were also noted, such as the presence of antiviral gene networks in a subset of the cells. Single-cell mRNA-seq data from 18 BMDCs showed that 15 cells strongly expressed antiviral and inflammatory defense cytokines, while the remaining three cells showed lower levels of expression. Different maturity states explain part of the variation, although the differences were also identified in the regulatory circuit-associated gene expression suggesting that a more intrinsic difference in cell phenotypes was present.

Vollmers et al. "Novel exons and splice variants in the human antibody heavy chain identified by single cell and single molecule sequencing." *PLoS One* (2015).

B cell receptors (membrane-bound Ig) and the corresponding secreted antibodies are created from recombinant processes that introduce high levels of variability within the B cell repertoire. Using single-cell C1 mRNA-seq with validation by bulk long-read sequencing, Vollmers et al. discovered novel exon expression in the heavy chain regions of the membrane-bound form of IgA and IgM in a minority (<1%) of B cells. Isoforms in the heavy chain sequence were also identified at the single-cell level with other membrane-bound isotypes, but were found only in rare cells. The functional difference of the cells expressing these isoforms remains to be determined. The isoforms within a given isotype (IgG, IgM) are still the subject of discovery, as are the functional differences the isoforms impart.

Key takeaways

- Full-length transcript mRNA-seq data contains not only the genes that are expressed but also specific isoform content.
- Cells may express a single isoform rather than a stochastic mixture. Isoform preference at the cell-specific level is likely correlated to specific cell functions.
- The functional implications for different isoforms highlight the need to correlate a given isoform with other gene expression biomarkers at the single-cell level.

Conclusion

The application of single-cell genomics techniques to the field of immunology has enabled deep profiling of the highly heterogeneous cells of the immune system and has led to great

advances in our understanding of its complexity. The papers overviewed herein provide a snapshot of these applications and how the sensitivity and data quality provided by the C1 system have been crucial to this success.

The ability to reconstruct paired TCR α and β chain sequences from single-cell mRNA-seq data has provided a way to link T cell phenotype and function to clonotype, opening the door to new discoveries and potential treatment strategies in the areas of autoimmunity, cancer research and immunotherapy.

The use of qPCR and mRNA-seq at the single-cell level has also enabled the identification of novel cell types and functions by providing an unprecedented level of resolution when studying gene expression and regulatory networks. Such single-cell resolution is especially important in understanding highly heterogeneous immune responses, how immune cells transition from one state to the next and the control of splice variation. Recent advances in our ability to harness the immune system to fight disease highlight the importance of basic research to more fully understand and characterize the mechanisms underlying immune cell function and subtypes.

Citations

Deshpande, M., Venuprasad, K., Parab, P.B. et al. "A novel CD28 mRNA variant and simultaneous presence of various CD28 mRNA isoforms in human T lymphocytes." *Human Immunology* 63 (2002): 20–23.

Gaublomme, J.T., Yosef, N., Lee, Y. et al. "Single-cell genomics unveils critical regulators of Th17 cell pathogenicity." *Cell* 163 (2015): 1,400–1,412.

Joly, A., Liu, S., Dahlberg, C.I.M. et al. "Foxp3 lacking exons 2 and 7 is unable to confer suppressive ability to regulatory T cells *in vivo*." *Journal of Autoimmunity* 63 (2015): 23–30.

Mahata, B., Zhang, X., Kolodziejczyk, A.A. et al. "Single-cell RNA sequencing reveals T helper cells synthesizing steroids *de novo* to contribute to immune homeostasis." *Cell Reports* 7 (2014): 1,130–1,142.

Proserpio, V., Piccolo, A., Haim-Vilmovsky, L. al. "Single-cell analysis of CD4+ T-cell differentiation reveals three major cell states and progressive acceleration of proliferation." *Genome Biology* 17(103) (2016): doi:10.1186/s13059-016-0957-5.

Shalek, A.K., Satija, R., Adiconis, X. et al. "Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells." *Nature* 498 (2013): 236–240.

Stubbington, M.J.T., Lonnberg, T., Proserpio, V. et al. "T cell fate and clonality inference from single-cell transcriptomes." *Nature Methods* 13 (2016): 329–332.

Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A. et al. "A census of human transcription factors: function, expression and evolution." *Nature Reviews Genetics* 10.4 (2009): 252-263.

Vollmers, C., Penland, L., Kanbar, J.N., Quake, S.R. "Novel exons and splice variants in the human antibody heavy chain identified by single cell and single molecule sequencing." *PLoS One* 10(1) (2015): e0117050.

Ward, F.J., Dahal, L.N., Khanolkar, R.C. et al. "Targeting the alternatively spliced soluble isoform of CTLA-4: prospects for immunotherapy?" *Immunotherapy* 6 (2014): 1,073–1,084.

CORPORATE HEADQUARTERS
7000 Shoreline Court, Suite 100
South San Francisco, CA 94080 USA
Toll-free: 866 359 4354 in the US
Fax: 650 871 7152
fluidigm.com

SALES
North America | +1 650 266 6170 | info-us@fluidigm.com
Europe/EMEA | +33 1 60 92 42 40 | info-europe@fluidigm.com
China (excluding Hong Kong) | +86 21 3255 8368 | info-china@fluidigm.com
Japan | +81 3 3662 2150 | info-japan@fluidigm.com
All other Asian countries | +1 650 266 6000 | info-asia@fluidigm.com
Latin America | +1 650 266 6000 | info-latinamerica@fluidigm.com